



# **STIC Search Report**

## **Biotech-Chem Library**

STIC Database Tracking Number: 1022784

**To: Sheridan Swope**  
**Location: CM1-10D01**  
**Art Unit: 1652**  
**Wednesday, August 27, 2003**

**Case Serial Number: 09/696872**

**From: Beverly Shears**  
**Location: Biotech-Chem Library**  
**CM1-1E05**  
**Phone: 308-4994**

**beverly.shears@uspto.gov**

### **Search Notes**

**This Page Blank (uspto)**

**From:** Swope, Sheridan  
**Sent:** Monday, August 25, 2003 6:21 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** FW: 09696872

-----Original Message-----

**From:** Chan, Christina  
**Sent:** Monday, August 25, 2003 6:20 PM  
**To:** Swope, Sheridan; STIC-Biotech/ChemLib  
**Subject:** RE: 09696872

**Please rush. Thanks Chris**

*Chris Chan*

TC 1600 New Hire Training Coordinator and SPE 1644  
308-3973  
CM-1, 9B19

-----Original Message-----

**From:** Swope, Sheridan  
**Sent:** Monday, August 25, 2003 5:18 PM  
**To:** Chan, Christina  
**Subject:** 09696872

May I have this rushed?

For 09696872, pls search:

SID 23 against the AA and NT data bases.

SID 24 against the AA and NT data bases.

SID 37 against the AA and NT data bases.

Sheridan Swope, Ph.D.  
Patent Examiner, AU 1652  
Recombinant Enzymes  
sheridan.swope@uspto.gov  
703-305-1696 (voice)  
703-308-3014 (FAX)  
Mailbox: CM1 Rm10D01  
Office: CM1 Rm12D12

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**This Page Blank (uspto)**



# SEARCH REQUEST FORM

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

## STAFF USE ONLY

Date completed: 08-27-03  
Searcher: Beverly 24994  
Terminal time: 20  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 25  
Number of Searches: \_\_\_\_\_  
Number of Databases: \_\_\_\_\_

### Search Site

\_\_\_\_\_ STIC  
\_\_\_\_\_ CM-1  
\_\_\_\_\_ Pre-S

### Type of Search

\_\_\_\_\_ N.A. Sequence  
\_\_\_\_\_ A.A. Sequence  
\_\_\_\_\_ Structure  
\_\_\_\_\_ Bibliographic

### Vendors

\_\_\_\_\_ IG  
\_\_\_\_\_ STN  
\_\_\_\_\_ Dialog  
\_\_\_\_\_ APS  
\_\_\_\_\_ Geninfo  
\_\_\_\_\_ SDC  
\_\_\_\_\_ DARC/Questel  
☒ Other CGN

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 26, 2003, 16:08:15 / Search time 4834.58 Seconds  
(without alignments)  
922.344 Million cell updates/sec

Title: US-09-696-872-23  
Sequence: 1 MRVMTGLALAAVCSAAK.....PKPQKPEPEGTGSSEKDEL 109

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Fgapop 10.0 , Fgapext 0.5  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USFO.spool/US09696872/runat\_26082003\_151137\_3223/app\_query.fasta.1.462  
-DB=genEmbl -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCH=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=spct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US09696872 @CGN 1 1 3608 @runat\_26082003\_151137\_3223 -NCPU=6 -ICPU=3  
-NO MAP -LARGOUIRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vir:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	579	99.1	372	6	AR121628 Sequence
2	459	78.6	372	6	AR121631 Sequence
3	429	73.5	372	6	AR121630 Sequence
4	428.5	73.4	387	6	AR121624 Sequence
5	422	72.3	444	6	AR121632 Sequence
6	414.5	71.0	387	6	AR121623 Sequence
7	371.5	63.6	369	6	AR121627 Sequence
8	345.5	59.2	369	6	AR121626 Sequence
9	337.5	57.8	357	6	AR121625 Sequence
10	304.5	52.1	315	6	AR121629 Sequence
11	262.5	44.9	2274	6	AB086984 Homo sapi
12	262.5	44.9	2439	6	AX329932 Sequence
13	262.5	44.9	2439	6	AX330698 Sequence
14	262.5	44.9	2439	6	AX330917 Sequence
15	262.5	44.9	2439	6	HUMCOMP
16	266.5	43.9	764	6	BD124595
17	266.5	43.9	764	6	BD126165
18	256.5	43.9	1779	6	BD127132
19	256.5	43.9	1779	6	AK074508
20	247.5	42.4	2302	4	AF325902
21	245	42.0	305	12	EVE291687
22	236	40.4	2438	10	AF033530
23	234	40.1	2421	10	RNCOMP
24	227.5	39.0	2710	9	BC033676
25	210	36.0	323	12	EVE291685
26	155	26.5	302	12	EVE291686
27	154.5	26.5	46275	9	AC0003107
28	153	26.2	4469	9	AB018333
29	153	26.2	4744	9	AK025495
30	153	26.2	154288	9	AL513164
31	153	26.2	170573	2	AC138890
32	149	25.5	144	6	A39800
33	149	25.5	177	6	A39798
34	146.5	25.1	3160	10	AF102887
35	146.5	25.1	3365	9	BC050456
36	146	25.0	8724	4	CDR131945
37	145.5	24.9	3074	6	AX094825
38	145.5	24.9	3074	6	AX336954
39	145.5	24.9	3074	6	AX480839
40	145.5	24.9	3074	9	HSTR0M4
41	144.5	24.7	3123	10	RNTSP4
42	144	24.3	78	6	A92276
43	142	24.3	85877	2	AC091343
44	142	24.3	244328	7	AC094938
45	141	24.1	210148	10	AC122864

RESULT 1

## ALIGNMENTS

AR121628 LOCUS AR121628 372 bp DNA linear PAT 16-MAY-2001  
 DEFINITION Sequence 24 from patent US 6160088.  
 ACCESSION AR121628  
 VERSION AR121628.1 GI:14105204  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 372)  
 AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.  
 TITLE KDEL receptor inhibitors  
 JOURNAL Patent: US 6160088-A 24 12-DEC-2000;  
 FEATURES  
 source 1..372  
 Location/Qualifiers  
 BASE COUNT 91 a 102 c 122 g 57 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4,476-45 Length: 372  
 Score: 579.00 Matches: 108  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 99.14% Indels: 0  
 Gaps: 0  
 DB:

US-09-696-872-23 (1-109) x AR121628 (1-372)

OY 2 ArgTYrMetIleLeuGlyLeuLeuAlaLeuAlaValaCySerAlaAlaIleYsgly 21  
 DB 16 AGGTACATGATTTTATGGCTTGTCTCCCTTGGCGAGCTCTGACGCGTCCAAAAAGGA 75  
 OY 22 SerSerLeuGlyGlyAspCySeSerAspLeuGlyProGlnMetLeuArgGluLeuGln 41  
 DB 76 TCCAGCCTGGGTGGAGACTGTTGTTCAGACCTGGGCGCGAGATCTTGGGAACTGCG 135  
 OY 42 GluThrAsnAlaAlaLeuGlnAspValaArgAspTrpLeuArgGlnValaArgGluIle 61  
 DB 136 GAACCAACGCGGCGCTGAGAGACGTGGGAGCTGGCTGCGGACAGGTCAAGGATC 195  
 OY 62 ThrPheLeuIysAsnThrValMetGluCyAspAlaCyGlyProGlnProGlnProlys 81  
 DB 196 ACGTTCCTGAATAAACACCGGTGATGAGTGTGACGCGTGGGCGCGCAGCGCGA 255  
 OY 82 ProGlnProGlnProGlnProGlnProlysPProGlnProGlnProGlnGlyThr 101  
 DB 256 CCGACGCGGAGCGCCGACGCGGAAACCGGACCGGAAACCGGAAAGGTACC 315  
 OY 102 GlySerSerGlyIysAspGluLeu 109  
 DB 316 GGATCATCAGAAAAAGATGAGTTG 339

RESULT 2  
 AR121631 LOCUS AR121631 372 bp DNA linear PAT 16-MAY-2001  
 DEFINITION Sequence 30 from patent US 6160088.  
 ACCESSION AR121631  
 VERSION AR121631.1 GI:14105207  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 372)  
 AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.  
 TITLE KDEL receptor inhibitors  
 JOURNAL Patent: US 6160088-A 30 12-DEC-2000;  
 FEATURES  
 source 1..372  
 Location/Qualifiers  
 BASE COUNT 102 a 96 c 106 g 68 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5,296-34 Length: 372  
 Score: 459.00 Matches: 84  
 Percent Similarity: 86.11% Conservative: 9  
 Best Local Similarity: 77.78% Mismatches: 15  
 Query Match: 78.60% Indels: 0  
 Gaps: 0  
 DB:

US-09-696-872-23 (1-109) x AR121631 (1-372)

OY 2 ArgTYrMetIleLeuGlyLeuLeuAlaLeuAlaValaCySerAlaAlaIleYsgly 21  
 DB 16 AGGTACATGATTTTATGGCTTGTCTCCCTTGGCGAGCTCTGACGCGTCCAAAAAGGA 75  
 OY 22 SerSerLeuGlyGlyAspCySeSerAspLeuGlyProGlnMetLeuArgGluLeuGln 41  
 DB 76 TCCAGCCTGGGTGGAGACTGTTGTGGGACTTTAACCGGACGTTCTTGGGCAATGACA 135  
 OY 42 GluThrAsnAlaAlaLeuGlnAspValaArgAspTrpLeuArgGlnValaArgGluIle 61  
 DB 136 CAATTAACCAACTCTGGAGAGAGTGAAGACCTTCTGAGACAGCAGGTTAAGAAACA 195  
 OY 62 ThrPheLeuIysAsnThrValMetGluCyAspAlaCyGlyProGlnProGlnProlys 81  
 DB 196 TCATTTTGGCGAAACACCATGATGTAATGCCAGGCTTGGCTCGGACGCGACCGGAAA 255  
 OY 82 ProGlnProGlnProGlnProGlnProlysPProGlnProGlnProGlnGlyThr 101  
 DB 256 CCGACGCGGAGCGCGGACCGGACCGGAAACCGGACCGGAAACCGGAAAGGTACC 315  
 OY 102 GlySerSerGlyIysAspGluLeu 109  
 DB 316 GGATCATCAGAAAAAGATGAGTTG 339

RESULT 3  
 AR121630 LOCUS AR121630 372 bp DNA linear PAT 16-MAY-2001  
 DEFINITION Sequence 28 from patent US 6160088.  
 ACCESSION AR121630  
 VERSION AR121630.1 GI:14105206  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 372)  
 AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.  
 TITLE KDEL receptor inhibitors  
 JOURNAL Patent: US 6160088-A 28 12-DEC-2000;  
 FEATURES  
 source 1..372  
 Location/Qualifiers  
 BASE COUNT 96 a 105 c 109 g 62 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3,16-31 Length: 372  
 Score: 429.00 Matches: 78  
 Percent Similarity: 84.26% Conservative: 13  
 Best Local Similarity: 72.22% Mismatches: 17  
 Query Match: 73.46% Indels: 0  
 Gaps: 0  
 DB:

US-09-696-872-23 (1-109) x AR121630 (1-372)

OY 2 ArgTYrMetIleLeuGlyLeuLeuAlaLeuAlaValaCySerAlaAlaIleYsgly 21  
 DB 16 AGGTACATGATTTTATGGCTTGTCTCCCTTGGCGAGCTCTGACGCGTCCAAAAAGGA 75  
 OY 22 SerSerLeuGlyGlyAspCySeSerAspLeuGlyProGlnMetLeuArgGluLeuGln 41  
 DB 76 TCCAGCCTGGGTGGAGACTGTTGTGGGAGTGAAGACCTTCTGAGACAGCAGGTTAAGAAACA 135  
 OY 42 GluThrAsnAlaAlaLeuGlnAspValaArgAspTrpLeuArgGlnValaArgGluIle 61

Db	Sequence	Length	Score	Alignment	Score	Length	Score	Alignment
Db	136 TCCTTCACACGATCTCTAGTGGAGCTTGGGAGCAGCATCCGAGACCAAGTGAAGAAATG	195						
QY	62 ThrPheLeuLysValSerThrValMetGluCysAspAlaCysGlyProGlnProGlnProLys	81						
Db	196 TCACATCATCCGGAACACCATCATGAGAGTGTGAGGTGTGGTCCGACCCGACCGGAAA	255						
QY	82 ProGlnProGlnProGlnProGlnProGlnProLysProGlnProGlnProGlnProGln	101						
Db	256 CCGCAGCCGCGACGCCGACCCGACCCGAGAAACCGACGCGGAAACCGGAAAGGTACC	315						
QY	102 GlySerSerGluLysAspGluLeu	109						
Db	316 GGATCATCATGAAAAAGATGAGTTG	339						
RESULT 4	AR121624	387 bp	DNA	Linear	PAT 16-MAY-2001			
LOCUS	Sequence 16 from patent US 6160088.							
ACCESSION	AR121624							
VERSION	AR121624.1							
KEYWORDS	GI:14105200							
SOURCE	Unknown.							
ORGANISM	Unknown.							
REFERENCE	1 (bases 1 to 387)							
AUTHORS	Rochman, J. E., Mayhew, M. and Hoe, M. H.							
TITLE	KDEL receptor inhibitors							
JOURNAL	Patent: US 6160088-A 16 12-DEC-2000;							
FEATURES	Location/Qualifiers							
source	1..387							
	./organism="unknown"							
BASE COUNT	93 a 107 c 126 g 61 t							
ORIGIN								
Alignment Scores:								
Pred. No.:	3.6e-31	Length:	387					
Score:	428.50	Matches:	87					
Percent Similarity:	80.36%	Conservative:	3					
Best Local Similarity:	77.68%	Mismatches:	5					
Query Match:	73.37%	Indels:	17					
DB:	6	Gaps:	3					
US-09-696-872-23 (1-109) x AR121624 (1-387)								
QY	8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysGlySerSerLeuGlyGlyAsp	27						
Db	40 TTGCTGCTGCTGGGGCGCGTG-----CGGCCGAGGATCCAGCTGGGTGGAAC	90						
QY	28 CysCysSerAspLeuGlyProGlnMetLeuLysGluLeuGlnGluThrAsnAlaLeu	47						
Db	91 TGTGCT-----CCACAGATGCTTCGAGAACTCCAGGAGACTAATGGCGCGCTG	138						
QY	48 GlnAspValArgAspTrpLeuArgGlnIleValArgGluIleThrPheLeuLysAsnThr	67						
Db	139 CAAGACGGAAGAGCTCTTGGACAGCAGGTCAAGAGATCACCTTCTGAAGAAATCG	198						
QY	68 ValMetGlnCysAspAlaCysGly-----ProGln	77						
Db	199 GTGATGGAATGAGCTTGGCGAATGACGCCGACGACCCCGGATCAATGTCGAG	258						
QY	78 ProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProLysProGln	97						
Db	259 CCGCAGCCGGAACCGCAGCCGACGCCGACGCCGAGAAACCGAACCGGAAACCGGAA	318						
QY	98 ProGlnGlyThrGlySerSerSerGluLysAspGluLeu	109						
Db	319 CCGAAGGTACCGGATCATCGAAAAAATGAGTTG	354						
RESULT 5	AR121632	444 bp	DNA	Linear	PAT 16-MAY-2001			
LOCUS	Sequence 35 from patent US 6160088.							
DEFINITION								

ACCESSION	AR121632				
VERSION	AR121632.1	GI:14105208			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 444)				
TITLE	Rothman,J.E., Mayhew,M. and Hoe,M.H.				
JOURNAL	KDEL receptor inhibitors				
FEATURES	Patent: US 6160088-A 35 12-DEC-2000;				
source	Location/Qualifiers				
	1..444				
BASE COUNT	/organism="unknown"				
ORIGIN	115 a 122 c 134 g 73 t				
<hr/>					
Alignment Scores:					
Pred. No.:	1..65e-30	Length:	444		
Score:	422.00	Matches:	82		
Percent Similarity:	84.85%	Conservative:	2		
Best Local Similarity:	82.83%	Mismatches:	1		
Query Match:	72.26%	Indels:	14		
DB:	6	Gaps:	2		
<hr/>					
US-09-696-872-23 (1-109) x AR121632 (1-444)					
QY	21 GlySerSerIeuGlyGlyAspCyS CySeSerAspleuGlyProglInMetLeuArgGluleu	40			
Db	127 GGATCCAGCGCTGGGTGAAGA CTGTGT-----CCACAGATGCTTCGAAACTC	174			
QY	41 GlnGluThrAsnAlaAlaLeuGlnAspValAlrAspTrpleuArgGlnGlnValArgGlu	60			
Db	175 CAGAGAGCTAATGGGGCGCTGCAGAGAGTGAGAGCTCTTGCACAGCACAGTCAAGAG	234			
QY	61 lIethrheleuLySaenThrrvalMetGlucysAspAlaCysgLy-----	75			
Db	235 ATCACCTTCTCGAAGATACCGGTATGGAATGTACCTTCCGGAATGCAGCCCCGACGC	294			
QY	76 -----ProglInProglInProlyBProglInProglInProglInProglInPro	90			
Db	295 ACCCGCGTAGTAAGTCCGACGCCGACCGCAAACCGACGCCGACGCCGACGCCG	354			
QY	91 LyBProglInProlyBProglInProglInGlyThrGlySerSerGlyLysBGluleu	109			
Db	355 AAACCGACGCCGAACCGGAACCGGAAGTAGTCCGATCATCAGAAAAAGATGAGTTG	411			
<hr/>					
RESULT 6					
LOCUS	AR121623	387 bp	DNA	linear	PAT 16-MAY-2001
DEFINITION	Sequence 14 from Patent US 6160088.				
ACCESSION	AR121623				
VERSION	AR121623.1	GI:14105199			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 387)				
AUTHORS	Rothman,J.E., Mayhew,M. and Hoe,M.H.				
TITLE	KDEL receptor inhibitors				
JOURNAL	Patent: US 6160088-A 14 12-DEC-2000;				
FEATURES	Location/Qualifiers				
source	1..387				
BASE COUNT	/organism="unknown"				
ORIGIN	94 a 110 c 125 g 58 t				
<hr/>					
Alignment Scores:					
Pred. No.:	7..05e-30	Length:	387		
Score:	414.50	Matches:	86		
Percent Similarity:	79.46%	Conservative:	3		
Best Local Similarity:	76.79%	Mismatches:	6		
Query Match:	70.98%	Indels:	17		
DB:	6	Gaps:	3		

US-09-696-872-23 (1-109) x AR121623 (1-387)

QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaAlaLysGlySerSerLeuGlyGlyAsp 27  
 Db 40 TTGCTGCTGCTGGGGCGCGTG-----CGGGCCGAGAGATCCAGCTGGGTGA--- 87

QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47  
 Db 88 -----GACCTAGCCCAACAGATGCTTCGAGAACTCCAGAGACATCATCGCGCGCTG 138

QY 48 GlnAspValaArgAspTrpLeuArgGlnGlnValaArgGluIleThrPheLeuLysAsnThr 67  
 Db 139 CAAGACGAGAGAGAGCTCTTGAGACAGCAGTCAGAGATCATCCTCTTGAAAGATACG 198

QY 68 ValMetGluCysAspAlaCysGly-----ProGln 77  
 Db 199 GTGATGAAATGTGAGCTTGGCGAATGCAGCCCGCAGCCGACCCCGGTAATGTCGCGAG 258

QY 78 ProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProLysProGlu 97  
 Db 259 CCGGACCGGAAACCGCAGCCGCGCAGCCGCGAGCCGAAACCGCAGCCGAAACCGGAA 318

QY 98 ProGluGlyThrGlySerSerGlyLysAspGluLeu 109  
 Db 319 CCGAAGGTACCGGATCATCAGAAAAAGATGAGTTG 354

RESULT 7  
 AR121627 369 bp DNA linear PAT 16-MAY-2001  
 LOCUS AR121627 Sequence 22 from patent US 6160088.  
 DEFINITION AR121627  
 ACCESSION AR121627 GI:14105203  
 VERSION AR121627.1  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 369)  
 AUTHORS Rochman,J.E., Mayhew,M. and Hoe,M.H.  
 TITLE KDEL receptor inhibitors  
 JOURNAL Patent: US 6160088-A 22 12-DEC-2000;  
 FEATURES  
 source 1..369  
 /organism="unknown"  
 BASE COUNT 94 a 94 c 121 g 60 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 6.23e-26 Length: 369  
 Score: 371.50 Matches: 68  
 Percent Similarity: 79.41% Conservative: 13  
 Best Local Similarity: 66.67% Mismatches: 18  
 Query Match: 63.61% Indels: 3  
 Gaps: 1

US-09-696-872-23 (1-109) x AR121627 (1-369)

QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaAlaLysGlySerSerLeuGlyGlyAsp 27  
 Db 40 TTGCTGCTGCTGGGGCGCGTG-----CGGGCCGAGAGATCCAGCTGGGTGA--- 90

QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47  
 Db 91 TGTGTGGTGGAGTCAGACAGACAGTTCGTCAGATTAACCAATGATCATGATGCTG 150

QY 48 GlnAspValaArgAspTrpLeuArgGlnGlnValaArgGluIleThrPheLeuLysAsnThr 67  
 Db 151 GGAGGCTCCGAGATGTCAGTGAACAGCAGGTCGAAAGAACCATGTTCTTGAAACACC 210

QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProLysProGlnProGlnProGlu 87  
 Db 211 ATTGCAAGATGCGAGGCTGTGGCCCGCAGCCGCGAAACCGCAGCCGAGCCGAG 270

QY 88 ProGlnProLysProGlnProLysProGluProGluGlyThrGlySerSerGlyLysAsp 107  
 Db 271 CCGCAGCCGAAACCGCAGCCGAAACCGGAAACCGGAAAGTACCGGATCATCAGAAAAAGAT 330

QY 108 GlnLeu 109  
 Db 331 GAGTTG 336

RESULT 8  
 AR121626 369 bp DNA linear PAT 16-MAY-2001  
 LOCUS AR121626 Sequence 20 from patent US 6160088.  
 DEFINITION AR121626  
 ACCESSION AR121626 GI:14105202  
 VERSION AR121626.1  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 369)  
 AUTHORS Rochman,J.E., Mayhew,M. and Hoe,M.H.  
 TITLE KDEL receptor inhibitors  
 JOURNAL Patent: US 6160088-A 20 12-DEC-2000;  
 FEATURES  
 source 1..369  
 /organism="unknown"  
 BASE COUNT 87 a 103 c 120 g 59 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.56e-23 Length: 369  
 Score: 345.50 Matches: 64  
 Percent Similarity: 76.47% Conservative: 14  
 Best Local Similarity: 62.75% Mismatches: 21  
 Query Match: 59.16% Indels: 3  
 Gaps: 1

US-09-696-872-23 (1-109) x AR121626 (1-369)

QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaAlaLysGlySerSerLeuGlyGlyAsp 27  
 Db 40 TTGCTGCTGCTGGGGCGCGTG-----CGGGCCGAGAGATCCAGCTGGGTGA--- 90

QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47  
 Db 91 TGTGTGGGAGACAGACCAAGCATTCGTCACCCAGCTCACCTCTTCACACAGATCCTA 150

QY 48 GlnAspValaArgAspTrpLeuArgGlnGlnValaArgGluIleThrPheLeuLysAsnThr 67  
 Db 151 GTGAGCTTCGGAGACATCCAGACAGGTCGAAAGAAATGTCATCATCCGAAACACC 210

QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProLysProGlnProGlnProGlu 87  
 Db 211 ATCATGAGTTCAGAGTGTGCTGCTCCGAGCCGAGCCGAAACCGCAGCCGAGCCGAG 270

QY 88 ProGlnProLysProGlnProLysProGluProGluGlyThrGlySerSerGlyLysAsp 107  
 Db 271 CCGCAGCCGAAACCGCAGCCGAAACCGGAAACCGGAAAGTACCGGATCATCAGAAAAAGAT 330

QY 108 GlnLeu 109  
 Db 331 GAGTTG 336

RESULT 9  
 AR121625 357 bp DNA linear PAT 16-MAY-2001  
 LOCUS AR121625 Sequence 18 from patent US 6160088.  
 DEFINITION AR121625  
 ACCESSION AR121625 GI:14105201  
 VERSION AR121625.1  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 357)

AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.  
TITLE KDEL receptor inhibitors  
JOURNAL Patent: US 6160088-A 18 12-DEC-2000;  
FEATURES Location/Qualifiers  
source 1..357  
/organism="unknown"  
BASE COUNT 84 a 100 c 114 g 59 t  
ORIGIN

Alignment Scores:

Pred. No.: 8.25e-23 Length: 357  
Score: 337.50 Matches: 65  
Percent Similarity: 75.49% Conservative: 12  
Best Local Similarity: 63.73% Mismatches: 18  
Query Match: 57.79% Indels: 7  
DB: 6 Gaps: 2

US-09-696-872-23 (1-109) x AR121625 (1-357)

Qy 8 LeuLeuAlaLeuAlaAlaValCySerAlaAlaValAlaValSerLeuGlyGlyAsp 27  
Db 40 TTGCTGCTGCTGGGCGCGCTG-----CGGGCCGAGGATCCAGCTGGTGAGAC 90  
Qy 28 CySerAspLeuGlyProGlnMetLeuArgLeuGlnGluThrAsnAlaAlaLeu 47  
Db 91 TGTGTAAAGGCAATTG-----GTCAACCACTCACCCTCTTCAACCACTCCTA 138  
Qy 48 GlnAspValaGAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLeuAsnThr 67  
Db 139 GTGAGGCTTCGGGAGCAGATCCGAGACCAAGTGAAGAAATGTCATCATCCGGAACCC 198  
Qy 68 ValMetGluCyAspAlaCyGlyProGlnProGlnProGlnProGlnProGln 87  
Db 199 ATCATGAGTTCAGGTGTGGGTCCGAGCCGAGCCGAAACCGCAGCCGAGCCGAG 258  
Qy 88 ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 107  
Db 259 CCGAGCGCGAAACCGCAGCCGAAACCGGAAACCGGAAAGTACCGATCATCAGAAAAAGAT 318  
Qy 108 GluLeu 109  
Db 319 GAGTTG 324

RESULT 10 AR121629 315 bp DNA linear PAT 16-MAY-2001  
LOCUS AR121629  
DEFINITION Sequence 26 from patent US 6160088.  
ACCESSION AR121629  
VERSION AR121629.1 GI:14105205  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 315)  
AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.  
TITLE KDEL receptor inhibitors  
JOURNAL Patent: US 6160088-A 26 12-DEC-2000;  
FEATURES Location/Qualifiers  
source 1..315  
/organism="unknown"  
BASE COUNT 81 a 85 c 79 g 70 t  
ORIGIN

Alignment Scores:

Pred. No.: 8.03e-20 Length: 315  
Score: 304.50 Matches: 70  
Percent Similarity: 66.67% Conservative: 14  
Best Local Similarity: 63.06% Mismatches: 13  
Query Match: 52.14% Indels: 25  
DB: 6 Gaps: 2

US-09-696-872-23 (1-109) x AR121629 (1-315)

Qy 2 ArgTrpMetIleLeuGlyLeuLeuAlaAlaValCySerAlaAlaValAlaValSerLeuGly 21  
Db 16 AGTACATGATTTTAAAGCTTGCTGCTGCGGCACTTCGACGCGTGCACAAAAAGGA 75  
Qy 22 SerSerLeuGlyGlyAspCySerAspLeuGlyProGlnMetLeuArgGluLeuGln 41  
Db 76 TCCAGCTGGGTGGAGACTGTTGCACAAAGCTACA-GAATCTATTTCATATTTCTCTCT 134  
Qy 42 GluThrAsnAlaAlaLeuGlnAspVal-----ArgAsp-TrpLeuArgGlnGlnValAr 59  
Db 135 CATTTAATATGCTCTTGCTGATCTATCATCTGATGATGCTTCTC----- 180  
Qy 59 GlnIleThrPheLeuLysAsnThrValMetGluCyAspAlaCyGlyProGlnProGln 79  
Db 181 -----CGGAGCCGCA 191  
Qy 79 nProLysProGlnProGlnProGlnProGlnProGlnProLysProGlnProGln 99  
Db 192 GCCGAAACCCGAGCCGAGCCGAGCCGAGCCGAAACCCGAGCCGAAACCCGGA 251  
Qy 99 uGlyThrGlySerSerGlyLysAspGluLeu 109  
Db 252 AGTACCGGATCATCAGAAAAAGATGAGTTG 282

RESULT 11 AB086984 2274 bp mRNA linear PRI 07-JAN-2003  
AB086984  
LOCUS  
DEFINITION Homo sapiens comp mRNA for cartilage oligomeric matrix protein.  
ACCESSION AB086984  
VERSION AB086984.1 GI:27531065  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1  
AUTHORS Hashimoto,Y. and Mori,H.  
TITLE Human comp cDNA with 5 SNIPS  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2274)  
AUTHORS Hashimoto,Y. and Mori,H.  
TITLE Direct Submission  
JOURNAL Submitted (27-JUN-2002) Hiroshi Mori, Osaka City University Medical School, Neuroscience, 1-4-3 Asahimachi, Abeno-ku, Osaka 545-8585, Japan (E-mail: mori@med.osaka-cu.ac.jp, Tel: 81-6-6645-3920, Fax: 81-6-6645-3922)

FEATURES

source 1..2274  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/cell\_type="fibroblast"  
1..2274  
/gene="comp"  
1..2274  
/gene="comp"  
/codon\_start=1  
/product="cartilage oligomeric matrix protein"  
/protein\_id="BAC53888.1"  
/db\_xref="GI:27531066"  
/translation="MVPDTRACVLLTLTALGASGGGSPGSLGPMLEIETNA  
LQVPRMLRQVREITFLPKTWTVECDACNQGQSVRTGSPVRLHLHCAEFCFPGVAC  
LQTESGRCGCPAGFTGNSHCTDVECNHAFCFPRVRCINTSPGRFCACFPFGVSG  
PTIQGVGLAPAKANKQVCTDINCEFGQHNVCVNSVCINTRGSPGCGCPGFGVQDA  
SGCGAQRCPDPSBSBECHHADCVLERDGRSVCVGVMAQNGILCGRDTLDLDFP  
DEKLRCEPCQRDNCVTVNSGQEDVDRDGDADCDPDADGDPVNEKDCNCLVNP  
DQNTBEDKWDGACDNCRSQNDQDKDTDDGGRGACDDIDGDRIRNQNDCPRVNP  
SDKSDGSDGIGPACNCRPKSNPDQADVDHDPVGPACSDQDQDSDGHDSDNCT  
VPSAQSDSDGQDADCDDDNDNDQVPSRDNCRILVPNGQSDADRDGADGVCQDQF  
DADKLVNKLIDVCEMAEVLITDFRAQTIVLDEGSAQIDPNVNLNQGELIVQTNMS  
DPGLAVGYTAFNGVDFEGTFHVTVDVDDYAGFI FGYOSSSFYVVMKQMBDTYQA  
NPRFAVAEPGIGLKAVYSTGPGEQRLNMLMHTGDTESQVRLLMKDPRLNMGMDKXSY

gene  
CDS





Db	280	GCGGCCCTGCTCCA	294
RESULT 13			
LOCUS	AXJ30698	2439 bp	DNA
DEFINITION	Sequence 1207 from Patent WO0194629.	linear	PAT 09-JAN-2002
ACCESSION	AXJ30698		
VERSION	AXJ30698.1	GI:18103676	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE			
AUTHORS	1 Young, P.E., Augustus, M., Carter, K.C., Edner, R., Endress, G., Horrigan, S., Sopet, D.R. and Weaver, Z.		
TITLE	Cancer gene determination and therapeutic screening using signature gene sets		
JOURNAL	Patent: WO 0194629-A 1207.13-DEC-2001;		
FEATURES	Avalon Pharmaceuticals (US)		
source	Location/Qualifiers 1..2439		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	503 a 758 c 809 g 369 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	5,14e-15	Length:	2439
Score:	262.50	Matches:	58
Percent Similarity:	76.47%	Conservative:	7
Best Local Similarity:	68.24%	Mismatches:	16
Query Match:	44.95%	Indels:	5
DB:	6	Gaps:	1
US-09-696-872-23 (1-109) x AXJ30698 (1-2439)			
OY	8	LeuLeuAlaLeuAlaAlaValCySeSerAlaAlaIalylSerSerLeuGlYAsp	27
Db	53	CTGTGCACCCTGGTGCTGCCCTCGAGCCGTCGAGCAGAGGCGCAGAGCCGTTGGGC-----	106
OY	28	CysCYSerAepLeuGlYPProGImetLeuArgIuLeuGIngluThAsnAlaLeu	47
Db	107	-----TCAgAcCTGGGCCCGcAgATGCTTCGGGAATCGcAGAAcCAACGCGCGCTG	160
OY	48	GlNaSPvLaLrgASPTryLeuArgIuAlaAngSuIethRPhLeuLyAsnThr	67
Db	161	CAGGACGCGCGGACTGCTGGCGGAGcAGAGGcAGATcAGTCTTAATAAACCG	220
OY	68	VAlmetGlucyASpAlaCYeGlYPProGlnPProGlnPProGlnPProGln	87
Db	221	GTGATGAGAGTGTGACGCGCTGGCGGATGACAGCATAGTACGAC-CGgcCTTACCCAGGT	279
OY	88	ProGlnPProLYSPro	92
Db	280	GCGGCCCTGCTCCA	294
RESULT 14			
LOCUS	AXJ30917	2439 bp	DNA
DEFINITION	Sequence 1426 from Patent WO0194629.	linear	PAT 09-JAN-2002
ACCESSION	AXJ30917		
VERSION	AXJ30917.1	GI:18103896	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE			
AUTHORS	1 Young, P.E., Augustus, M., Carter, K.C., Edner, R., Endress, G., Horrigan, S., Sopet, D.R. and Weaver, Z.		
TITLE	Cancer gene determination and therapeutic screening using signatures		

JOURNAL	Gene sets
PATENT	Patent: WO 0194629-A 1426-13-DEC-2001;
AVALON	Avalon Pharmaceuticals (US)
FEATURES	Location/Qualifiers
SOURCE	1..2439 /organism="Homo sapiens" /mol_type="genomic DNA"
BASE COUNT	503 A     758 C     809 G     369 T
ORIGIN	/db_xref="taxon:9606"
Alignment Scores:	
Pred. No.:	5.14e-15 Length: 2439
Score:	262.50 Matches: 58
Percent Similarity:	76.47% Conservative: 7
Best Local Similarity:	68.24% Mismatched: 16
Query Match:	44.95% Indels: 5
DB:	Gaps: 1
US-09-696-872-23 (1-109) x AX330917 (1-2439)	
Dy	8 LeuleualaleualaalaValCysSerAlaAlalylserLeuglYasp 27
Dd	:::       53 CTGCTCACCTGGGTGCCCTCGAGCCGTCGGAACAGGCACCCTTGGG----- 106
Dy	28 CysCySerAspleuGlYProlImetLenuArgluIleunglnGuThraanaLaaleu 47
Dd	O7 -----TCAGACTCGGSCCGCAgATcCTTGCGAACTCAGAgaAACCAAcGCGCcCTG 160
Dy	48 GlNapValArGaAprTPleuaRglngInVaARguileTTPheloulyAAnthr 67
Dd	161 CAGAGCCTGGCGGACtVGctCGCGCaCAGtcACAgAtCACgtTCtgAAAAACAAG 220
Dy	68 ValMetLucCYsaAPALACYSGLYPrgINProginPrOlInPrOpGINProGLnPRogLn 87
Dd	:::       221 GTATTGAGtGTACgcGCTccGGAttCaCATgCatTAAGcAc -cgGCTTAcccAcGCT 279
Oy	88 ProGIInPrOLypSPRO 92
Dd	 280 GCggCCCCCTGCTCCA 294
RESULT_15	
HUMCOMP	2439 bp mRNA linear PRI 15-DEC-1994
LOCUS	Human germ-line oligomeric matrix protein (COMP) mRNA, complete cds.
DEFINITION	L32137
ACCESSION	L32137.1 GI:602449
VERSION	germline; matrix protein.
KEYWORDS	Homo sapiens (human)
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euarcharia; Primates; Catarrhini; Homniidae; Homo. 1 (bases 1 to 2439) Newton,G., Werenowicz,S., Morton,C.C., Copeland,N.G., Gilbert,D.J., Jenkins,N.A. and Lawler,J. Characterization of human and mouse cartilage oligomeric matrix protein Unpublished (1994)
TITLE	Original source text: Homo sapiens cartilage cDNA to mRNA.
REFERENCE	Authors
AUTHORS	
JOURNAL	
COMMENT	Location/Qualifiers
FEATRES	1..2439
SOURCE	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /map="19p13.1" /cell_type="chondrocyte" /issue_type="cartilage" /germline 1..2439 ./gene="COMP" 26..2299 ./gene="COMP" /standard_name="cartilage oligomeric matrix protein".
CDS	
gene	
CDS	

```
/note="putative"  
/codon_start=1  
/product="matrix protein"  
/protein_id="AA57253.1"  
/db_xref="GI:602450"  
/translation="MVPTACVLLITLALGASGGGSLGSDLGSPQMLREIQETNAALQADVADWLRQVREITFLKNTVMEDACMQDSVRTGLPSVRPLHCAFGCFPGVACIQTESGRCGPCPAGFTNGSHCTDNECNAPCFPRVCINTSPGRCBACPGYSGPTHQGVGLAFPAKANKOVCTDINECTGQHNVCVPSVNCINTRGSPQCGPQGVDOASGCCORGAORFCPDGSPBSCHEHADCLEIRGDSRSCVCRVMAANGIILGRTDLGGFDEKLRCPEPQCRKNCVTPVNSGQEDVDRDGDGDCPDADGCVPRKNCPLVRNPDKRTDERKMGDACNCRSPQKNDQKDTDQSGKADCDDDIDGKIRNQADNCPVENSQKSDSDGIGDACNCPQSNPQADVDDHFDGADCDSDQDQDGDHQSNDNCPPTVNSAQEDSDHDGQDADCDDDDDNDGVPDSRNCRLVNPQGEDADRDGVGVCDDFDADKVVNDKIDVCPENAEVTLTDFRAFTVLDPEGDAOIDPMMVVLNQGREGVCTMNSDPLAVGYTAFNGVDFEGTFHNTVTDDVAGFIFGYODSSFFVVMKQKQWQOANPFRVAVARPGIQKAVKSSGTGEGQLRNALMHTGPTESQVRLMKDPRNKGKDKSYRMFLQHRPQVGIRVRFYEGPELVADSNVLDITMRGRLGVFCPSQENITWNLRYRCNDTTPEDYETHQLRQA"
```

```
sig_peptide  
repeat_region  
repeat_region  
repeat_region  
polyA_signal  
polyA_site
```

```
BASE COUNT 503 a 758 c 809 g 369 t  
ORIGIN
```

```
Alignment Scores:  
Pred. No.: 5.14e-15 Length: 2439  
Score: 262.50 Matches: 58  
Percent Similarity: 76.47% Conservative: 7  
Best Local Similarity: 68.24% Mismatches: 16  
Query Match: 44.95% Indels: 5  
DB: 9 Gaps: 1
```

US-09-696-872-23 (1-109) x HUMCOMP (1-2439)

```
QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27  
DB 53 CTGCTACCCCTGGCTGGCTCGCGCGCTCCGACAGGCGCCAGAGCCCGTTGGG----- 106  
QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47  
DB 107 -----TCAGACCTGGGCCCGGACAGATGCTTCGGGAATGCCAGAAACCAACGCGCGCTG 160  
QY 48 GlnAspValArgAspTyrLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67  
DB 161 CAGGACGTCGCGGAGCTGGCTGGCGACAGCAGTCAGGAGATCAGTCTCTGAAAAACACG 220  
QY 68 ValMetGlnCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGlnProGln 87  
DB 221 GTGATGGAGTGTGACGCGTGGGATGCGACAGTCAGTACGAC-CGGCTTACCAAGCGCT 279  
QY 88 ProGlnProLysPro 92  
DB 280 GCGGCCCTGCTCCA 294
```

Search completed: August 26, 2003, 18:30:17  
Job time : 4837.58 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 26, 2003, 16:04:40 ; Search time 373.301 Seconds  
(without alignments) 788.208 Million cell updates/sec

Title: US-09-696-872-23  
Perfect score: 584  
Sequence: 1 MRYMIGLALAAVCSAAKK.....PKPQKPEPGTGSSEKDEL 109

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 segs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgm2\_1/USPTO.spool/US09696872/runat\_26082003\_151137\_3213/app.query.fasta\_1.462  
-DB=N-Geneseq\_19Jun03 -OPMT=factap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=plc -NDB=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US09696872.@CGN\_1.1.0.@runatc\_26082003\_151137\_3213 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEOUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N-Geneseq\_19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	584	100.0	369	AA250497	KDEL receptor inh1
2	464	79.5	369	AA250500	KDEL receptor inh1
3	434	74.3	369	AA250493	KDEL receptor inh1
4	428.5	73.4	387	AA250493	KDEL receptor inh1
5	422	72.3	387	AA250501	KDEL receptor inh1
6	414.5	71.0	387	AA250492	KDEL receptor inh1
7	371.5	63.6	369	AA250495	KDEL receptor inh1
8	345.5	59.2	369	AA250495	KDEL receptor inh1
9	337.5	57.8	357	AA250494	KDEL receptor inh1
10	309.5	53.0	312	AA250498	KDEL receptor inh1
11	262.5	44.9	755	AAA47734	Human COMP/TSP-1 c
12	262.5	44.9	925	AAA47735	Human COMP/TSP-2 c
13	262.5	44.9	2439	ABT10887	Human breast cancer
14	262.5	44.9	2439	ABT07751	Breast cancer-asso
15	262.5	44.9	2439	ABL62104	Colon adenocarcino
16	262.5	44.9	2439	ABL62870	Breast cancer rela
17	262.5	44.9	2439	ABL63089	Breast cancer rela
18	262.5	44.9	2439	ACC50112	Breast cancer asso
19	262.5	44.9	2439	ABX76334	Lung cancer-asso
20	265.5	43.9	764	AAK91566	Human CDNA 5'-end
21	266.5	43.9	764	AAK93136	Human CDNA clone r
22	266.5	43.9	1779	AAK94103	Human full-length
23	256.5	43.9	8532	ABA15407	Human nervous syst
24	154.5	26.5	8532	ABA15585	Human breast cance
25	154.5	26.5	46275	ABT10145	Human full-length
26	153	26.2	4843	AA544633	Human polynucleoti
27	153	26.2	4844	AAK51864	Human polynucleoti
28	153	26.2	4879	AAK52848	Human polynucleoti
29	149	25.5	144	AAO67370	VHC-AG-alpha-1 lin
30	149	25.5	177	AAO67368	VHC-AG-alpha-1 lin
31	145.5	24.9	2916	AAK52211	Human polynucleoti
32	145.5	24.9	3074	AAO66455	Human thrombospo
33	145.5	24.9	3074	AAO66455	Human thrombospo
34	145.5	24.9	3074	AAO66455	Human thrombospo
35	145.5	24.9	3074	AAO66455	Human thrombospo
36	142.5	24.4	1866	AAK53195	Human polynucleoti
37	138.5	23.7	320	AAK58415	CDNA #1091 encodin
38	138.5	23.7	2820	AAO66454	Xenopus thymospo
39	138	23.6	222	AAK71235	Cassette encoding
40	137	23.5	78	AAV27572	Nucleotide sequenc
41	133	22.8	900	AAK90746	DNA encoding the p
42	127	21.7	12290	ABK78918	E. coli CFT073 gen
43	126	21.6	96	AAK57135	Quadruplex DNA SEQ
44	126	21.6	1172	AAK23835	Sugarcane proline
45	126	21.6	11466	AAK3828	PyAc4-Ascl vector

## ALIGNMENTS

RESULT 1

AA250497 standard; DNA; 369 BP.

AA250497;

23-MAY-2000 (first entry)

KDEL receptor inhibitor-6 DNA.

KDEL receptor inhibitor; heat shock protein; immune response;

oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;

metastoma; carcinoma; glioblastoma; astrocytoma; oncogene;

infectious disease; allergy; autoimmune disease; ss.

Chimeric - Adenovirus E3.

Chimeric - Homo sapiens.

XX	Chimeric - Camelus sp.	Location/Qualifiers
XX	Key	10..339
XX	CDS	/*tag= a
XX	sig_peptide	/product= "KDEL receptor inhibitor protein"
XX	mat_peptide	10..69
XX		/*tag= b
XX		/note= "Derived from adenovirus E3"
XX		70..336
XX		/*tag= c
XX		/product= "Mature KDEL receptor inhibitor protein"
XX		97..234
XX		/*tag= d
XX		/note= "Human COMP pentamerisation domain"
XX		235..306
XX		/*tag= e
XX		/note= "Camel IgG linker domain"
XX	WC000006729-A1.	
XX	10-FEB-2000.	
XX	28-JUL-1999;	99WC-US17147.
XX	29-JUL-1998;	98US-0124671.
XX	(SLOK ) SLOAN KETTERING INST CANCER RES.	
XX	Rochman JE, Mayhew M, Hoe MH;	
XX	WPI: 2000-195296/17.	
XX	P-PSDB: AAY44963.	
XX	Inhibitors of the KDEL receptor which comprises an oligomerization	
XX	domain useful for promoting secretion of proteins which are normally	
XX	retained within the cell	
XX	Disclosure; Fig 6; 87pp; English.	
XX	The patent discloses the use of KDEL receptor inhibitor to promote	
XX	secretion of proteins that are normally retained within the cell such as	
XX	heat shock proteins by inhibiting KDEL receptor-mediated return of	
XX	protein complexes to endoplasmic reticulum. This makes the secreted heat	
XX	shock proteins more accessible to the immune system and improves immune	
XX	response to a target antigen. The inhibitor protein comprises several	
XX	subunits where each subunit comprises an oligomerisation domain and has	
XX	at its carboxy terminus a region which binds to a KDEL receptor. The	
XX	target antigen may be associated with diseases including neoplasia such	
XX	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and	
XX	astrocytoma, with defective tumour suppressor genes, oncogenes,	
XX	infectious diseases, allergy or autoimmune diseases. The present	
XX	sequence encodes KDEL receptor inhibitor comprising regions encoding a	
XX	cleavable signal peptide; the oligomerisation domain of human cartilage	
XX	oligomeric matrix protein (COMP) pentamerisation domain; a camel IgG	
XX	linker domain and the carboxy-terminal sequence KDEL. The subsequence	
XX	GDPC is an alteration of rat COMP which provides increased stability via	
XX	disulphide bonds.	
XX	Sequence 369 BP; 90 A; 102 C; 120 G; 57 T; 0 other;	
XX	Alignment Scores:	
XX	Pred. No.:	6.49e-45
XX	Score:	584.00
XX	Percent Similarity:	100.00%
XX	Best Local Similarity:	100.00%
XX	Query Match:	100.00%
XX	DB:	21
XX	Length:	369
XX	Matches:	109
XX	Conservative:	0
XX	Mismatches:	0
XX	Indels:	0
XX	Gaps:	0

Dd		10	ATAGGATACATGATTTTTACGGCTTGCTGCCCTTCGGCGAAGCTCAGCCGCTGCACAAAAA	69
Oy		21	GlySerSerLeuGlYglYAspCyCysSerAspIleuGlYProGlnMetLeuArgGlu	40
Dd		70	GGATCCAGCTCGGTGGAGACTGTGTTCAGACCCTGGGCCCCGACAGATGCTTCGGGAAC	129
Oy		41	GlnGlnThrAsnAlaIleLengInnApValArgAspTrpLeuApsGlnInnValArgGlu	60
Dd		130	CAGAAAACAACCGGCGCTTCAGAGACGTGCGGACTGCTGCGGACAGAGTCAGGGAG	189
Oy		61	IleThrPheLeuLySAshAnThrValMetGluCyAspAlaCySgIYProGlnProGlnPro	80
Dd		190	ATCACGCTTCCTGA AAAACACGGGTGATGAGAGTGTGACGCGGTCCGGGCGCAGCCG	249
Oy		81	LysProGlnIProGlnIProGlnIProGlnIProGlnIProGlnIProGlnIProGlnI	100
Dd		250	AAACCCAGCCGAGCGCGAGCGCGAGCGCGA AACCGCAACCGGAACCGGAAGGT	309
Oy		101	ThrGlySerSerGlnYAspGlnLeu	109
Dd		310	ACCGATCATCAGAAAAAGATGAGTTG	336
 RESULT 2 AAZ50500 ID AAZ50500 standard; DNA; 369 BP.				
Ac			AAZ50500;	
Xx				
Dt		23-MAY-2000	(first entry)	
Xx				
De		KDEL receptor inhibitor-9 DNA.		
Xx				
Kw		KDEL receptor inhibitor; heat shock protein; immune response;		
Kw		oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;		
Km		melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;		
Km		infectious disease; allergy; autoimmune disease; ss.		
Xx				
Os		Chimeric - Adenovirus E3.		
Os		Chimeric - Homo sapiens.		
Os		Chimeric - Camelus sp.		
Xx				
Fh		Key	Location/Qualifiers	
Ft		CDS	10..339	
Ft			/tag= a	
Ft			/product= "KDEL receptor inhibitor protein"	
Ft		sig_peptide	10..69	
Ft			/tag= b	
Ft			/note= "Derived from adenovirus E3"	
Ft		mat_peptide	70..336	
Ft			/tag= c	
Ft			/note= "Mature KDEL receptor inhibitor protein"	
Ft		misc_feature	97..234	
Ft			/tag= d	
Ft			/note= "Human TSP4 trimerisation domain"	
Ft		misc_feature	235..306	
Ft			/tag= e	
Ft			/note= "Camel IgG linker domain"	
Xx				
Pn		WO200006729-A1.		
Xx				
Pd		10-FEB-2000.		
Xx				
Pf		28-JUL-1999;	99MO-US17147.	
Xx				
Pf		29-JUL-1998;	98US-O124671.	
Xx				
Pa		(SLOK ) SLOAN KETTERING INST CANCER RES.		
Xx				
Ei		Rochman JE, Maynew M, Hoe MH;		
Xx				
Wp		wpi; 2000-195296/17.		
Dr		p-psdb; AAy44966.		
Xx				



```

US-09-696-872-23 (1-109) x AAZ50499 (1-369)

OY      1 MetarGTYrmEtIleuGIyleuLeuAlaleuAlaIaIaValCySSerAlaIaIySlyS 20
Db      10 ATGAGGTACATCATTTTAAAGGCTTGCTCGCCCTTGCGGACAGTCTGACAGCGCTGCCAAAAA 69
OY      21 GlySerSerIeuGIyGIyAspCySsSerAspIeuGIyProGIImetIeuArgGIuIeu 40
Db      70 GGATCCACCCCTGGGGGAGACTGTGTGGGGGAGCAGACCAAGGCAATGGTACCCAGCTC 129
OY      41 GInGIuTrIrrAsenAlaIaleuGIInAspValAraGAsPTrIleuArgGIInuValAraGIu 60
Db      130 ACCCTTCACACCAATACCTTACTAGAGCTTCGGGACGACATCCGAAACAGGTAAGAA 189
OY      61 ILeThrPhleuLeuYAsnThrValMetGIuCyAspAlaICySgIyProGIInProGIInPro 80
Db      190 ATGTACATCATCCGGAACACCATCATGTAGAGTGCAGGTGTGCAGTCCGACGCGGACCG 249
OY      81 LysProGIInProGIInProGIInProGIInProGIInProGIInProGIInProGIInGIu 100
Db      250 AAACCGGAGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGAA 309
OY      101 ThrGIySerSerGIuIyAspAlaIeu 109
Db      310 ACCGATCATCATCGAAAAAGATGAGTTG 336

RESULT 4
AAZ50493
ID      AAZ50493 standard; DNA; 387 BP.
AC      AAZ50493;
XX
XX      23-MAY-2000 (first entry)
DE
DE      KDEL receptor inhibitor-2 DNA.
XX
XX      KDEL receptor inhibitor; heat shock protein; immune response;
KW      oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
KW      melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
KW      infectious disease; allergy; autoimmune disease; ss.
XX
XX      Chimeric - Mus sp.
OS      Chimeric - Camelus sp.
OS      Chimeric - Rattus sp.
XX
XX      Key      Location/Qualifiers
FH      CDS      10..357
FT      /tag= a
FT      /product= "KDEL receptor inhibitor protein"
FT      sig_peptide 10..68
FT      /tag= b
FT      /note= "Derived from mouse Bip"
FT      mat_peptide 69..354
FT      /tag= c
FT      /product= "Mature KDEL receptor inhibitor protein"
FT      misc_feature 98..222
FT      /tag= d
FT      /note= "Rat COMP domain"
FT      misc_feature 253..324
FT      /tag= e
FT      /note= "Camel IgG linker domain"
XX
XX      WO200006729-A1.
XX
XX      10-FEB-2000.
XX
XX      28-JUL-1999; 99WO-US17147.
XX
XX      29-JUL-1998; 98US-0124671.
XX
XX      (SLOK ) SLOAN KETTERING INST CANCER RES.
XX

```

PI Roehman DE, Mayhew M, Hoe MH;  
XX WPI; 2000-195296/17.  
DR P-PSDB; AAY44959.  
XX  
XX  
PT Inhibitors of the KDEL receptor which comprises an oligomerization  
PT domain useful for promoting secretion of proteins which are normally  
PT retained within the cell -  
XX  
PS Disclosure; Fig 2; 87pp; English.  
XX  
XX  
CC The patent discloses the use of KDEL receptor inhibitor to promote  
CC secretion of proteins that are normally retained within the cell such as  
CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
CC shock proteins more accessible to the immune system and improves immune  
CC response to a target antigen. The inhibitor protein comprises several  
CC subunits where each subunit comprises an oligomerisation domain and has  
CC at its carboxy terminus a region which binds to a KDEL receptor. The  
CC target antigen may be associated with diseases including neoplasia such  
CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
CC infectious diseases, allergy or autoimmune diseases. The present  
CC sequence encodes KDEL receptor inhibitor comprising regions encoding a  
CC cleavable signal peptide; the oligomerisation domain from rat cartilage  
CC oligomeric matrix protein (COMP); a camel IgG linker domain and the  
CC carboxy-terminal sequence KDEL. The subsequence GDCC is an alteration of  
CC rat COMP which provides increased stability via disulphide bonds. This  
CC is introduced into host cells by suitable vectors.

SQ Sequence 387 BP; 93 A; 107 C; 126 G; 61 T; 0 other;

**Alignment Scores:**

Pred. No.:	1,059-30	length:	387
Score:	428.50	Matches:	87
Percent Similarity:	80.36%	Conservative:	3
Best local Similarity:	77.68%	Mismatches:	5
Query Match:	73.37%	Indels:	17
DB:	21	Gaps:	3

US-09-696-872-23 (1-109) x AA250493 (1-387)

QY 8 LeuLeuAlaLeuAlaAlaValCysSerLeuAlaAlaIysIysGlySerSerLeuGlyGlyAsp 27  
Db 40 TTCTCTCTGCTGGGCCGGGTG-----CGGGCCGGGATCCACCTGGGTGGAGAC 90  
QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgIuIeuGlnGlnIuTrpAlaAlaLeu 47  
Db 91 TGTGTG-----CCACAGATGCTTGAGAACTCCGAGGACTAAATGCGCGCTG 138  
QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGlnIleThrPheLeuIysAsnThr 67  
Db 139 CAAGACGTGAGACAGCTCTTGGCAGACAGAGTCAAGGAGATCACCTTCTTGMAGAAATACG 198  
QY 68 ValMetGlnCysAspAlaCysGly-----ProGln 77  
Db 199 GTATGGAAATGTGACGCTTGGCGAATGCAGCCGACGACACCCCGGAGTAAATGCCAG 258  
QY 78 ProGlnProIysProGlnProGlnProGlnProGlnProIysProGlnProIysProGln 97  
Db 259 CCGCAGCCGAAACCGCAGCCGACGCCGACGCCGACGCCGAAACCGCAGCCGAAACCGGAA 318  
QY 98 ProGlnGlyThrGlySerSerGluIysAspGluIeu 109  
Db 319 CCGAAGAGTACCGAGTACATCAAAAAAATAGATTG 354

RESULT	5
AAZ50501	
ID	AAZ50501 standard; DNA; 444 BP
XX	
AC	AAZ50501;
XX	
DT	23-MAY-2000 (first entry)

XX KDEL receptor inhibitor-10 DNA.  
 XX  
 DE  
 XX KDEL receptor inhibitor; heat shock protein; immune response;  
 KM oligomerization domain; neoplasia; sarcoma; lymphoma; leukaemia;  
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
 XX infectious disease; allergy; autoimmune disease; ss.  
 OS Chimeric - Mus sp.  
 OS Chimeric - Rattus sp.  
 OS Chimeric - Camelus sp.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 10..414  
 FT /tag= a  
 FT /product= "KDEL receptor inhibitor protein"  
 FT sig\_peptide 10..69  
 FT /tag= b  
 FT /note= "Derived from mouse Bip"  
 FT mat\_peptide 70..411  
 FT /tag= c  
 FT /product= "Mature KDEL receptor inhibitor protein"  
 FT misc\_feature 154..279  
 FT /tag= d  
 FT /note= "Rat COMP pentamerisation domain"  
 FT misc\_feature 310..381  
 FT /tag= e  
 FT /note= "Camel IgG linker domain"  
 XX  
 XX WO200006729-A1.  
 XX  
 PD 10-FEB-2000.  
 XX  
 XX 28-JUL-1999; 99WO-US17147.  
 XX  
 XX 29-JUL-1998; 98US-0124671.  
 XX  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 XX  
 XX Rothman JE, Mayhew M, Hoe MH;  
 DR WPI; 2000-195296/17.  
 DR P-PSDB; AAY44967.  
 XX  
 XX Inhibitors of the KDEL receptor which comprises an oligomerization  
 PT domain useful for promoting secretion of proteins which are normally  
 PT retained within the cell -  
 XX  
 PS Disclosure; Fig 10; 87pp; English.  
 XX  
 XX The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerisation domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a  
 CC cleavable signal peptide; a myc-tag; an N-glycosylation sequence; the  
 CC oligomerisation domain of rat cartilage oligomeric matrix protein  
 CC (COMP); a camel IgG linker domain and the carboxy-terminal sequence  
 CC KDEL. The subsequence GDCC is an alteration of rat COMP which provides  
 CC increased stability via disulphide bonds.  
 XX  
 XX Sequence 444 BP; 115 A; 122 C; 134 G; 73 T; 0 other;  
 SO  
 Alignment Scores: 4.79e-30 Length: 444  
 Pred. No.:

Score: 422.00 Matches: 82  
 Percent Similarity: 84.85% Conservative: 2  
 Best Local Similarity: 82.83% Mismatches: 1  
 Query Match: 72.26% Indels: 14  
 DB: 21 Gaps: 2  
 US-09-696-872-23 (1-109) x AA250501 (1-444)  
 QY 21 GlycerSerLeuGlyGIYAspCySeserAspLeuGlyProGlnMetLeuArgGluLeu 40  
 DB 127 GGATCCAGCTCGGTGGAGACTGTGT-----CCACGATGCTTCGAACTC 174  
 QY 41 GlnGluThrAsnAlaAlaLeuGlnAspValArgAspTrpLeuArgGlnValArgGlu 60  
 DB 175 CAGAGACTAATGCGCGCTCCAGACTGAGAGAGCTCTTGGCAGCAGGTACAAGAG 234  
 QY 61 IleThrPheLeuLysAsnThrValMetGluCyAspAlaCySgily----- 75  
 DB 235 ATCACCTTCTCGAAGAAATACGGTATGGAATGTGACCTTCGGGAATGCAGCCGCAGC 294  
 QY 76 -----ProGlnProGlnProLysProGlnProGlnProGlnPro 90  
 DB 295 ACCCCCGACTACTCCGACGCCGACCCGAAACCGAGCCGCGCAGCCGACGCG 354  
 QY 91 LysProGlnProLysProGlnProGlnGlyThrGlySerSerGlyLysAspGluLeu 109  
 DB 355 AAACCGACCGCGAACCAGAACCGGAAGTACCGGATCATCAGAAAAAGATGACTTG 411  
 RESULT 6  
 AA250492  
 ID AA250492 standard; DNA; 387 BP.  
 XX  
 XX AC AA250492;  
 XX  
 DT 23-MAY-2000 (first entry)  
 XX  
 XX KDEL receptor inhibitor-1 DNA.  
 XX  
 XX KDEL receptor inhibitor; heat shock protein; immune response;  
 KM oligomerization domain; neoplasia; sarcoma; lymphoma; leukaemia;  
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
 XX infectious disease; allergy; autoimmune disease; ss.  
 OS Chimeric - Mus sp.  
 OS Chimeric - Camelus sp.  
 OS Chimeric - Rattus sp.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 10..357  
 FT /tag= a  
 FT /product= "KDEL receptor inhibitor protein"  
 FT sig\_peptide 10..68  
 FT /tag= b  
 FT /note= "Derived from mouse Bip"  
 FT mat\_peptide 69..354  
 FT /tag= c  
 FT /product= "Mature KDEL receptor inhibitor protein"  
 FT misc\_feature 98..222  
 FT /tag= d  
 FT /note= "Rat COMP domain"  
 FT misc\_feature 253..324  
 FT /tag= e  
 FT /note= "Camel IgG linker domain"  
 XX  
 XX WO200006729-A1.  
 XX  
 PD 10-FEB-2000.  
 XX  
 XX 28-JUL-1999; 99WO-US17147.  
 XX  
 XX 29-JUL-1998; 98US-0124671.  
 XX  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 PA

XX Rothman JE, Mayhew M, Hoe MH;  
 XX WPI: 2000-195296/17.  
 DR P-PSDB; AA44958.  
 XX  
 XX Inhibitors of the KDEL receptor which comprises an oligomerisation  
 PT domain useful for promoting secretion of proteins which are normally  
 PT retained within the cell -  
 PS Disclosure: Fig 1, 87pp; English.  
 CC The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerisation domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a  
 CC cleavable signal peptide; the oligomerisation domain from rat cartilage  
 CC oligomeric matrix protein; a camel IgG linker domain and the carboxy  
 CC -terminal sequence KDEL. This is introduced into host cells by suitable  
 CC vectors.  
 XX  
 XX Sequence 387 BP; 94 A; 110 C; 125 G; 58 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1,996-29 Length: 387  
 Score: 414.50 Matches: 86  
 Percent Similarity: 79.468 Conservative: 3  
 Best Local Similarity: 76.794 Mismatches: 6  
 Query Match: 70.984 Indels: 17  
 DB: 21 Gaps: 3  
 US-09-696-872-23 (1-109) x AA250492 (1-387)

XX KDEL receptor inhibitor-5 DNA.  
 DE  
 XX  
 XX KDEL receptor inhibitor; heat shock protein; immune response;  
 KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;  
 KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
 KW infectious disease; allergy; autoimmune disease; ss.  
 XX  
 OS Chimeric - Mus sp.  
 OS Chimeric - Xenopus sp.  
 OS Chimeric - Camelus sp.  
 XX  
 XX Key Location/Qualifiers  
 FH CDS 10..339  
 FT /tag= a  
 FT /product= "KDEL receptor inhibitor protein"  
 FT sig\_peptide 10..68  
 FT /tag= b  
 FT /note= "Derived from mouse Bip"  
 FT mat\_peptide 69..336  
 FT /tag= c  
 FT /product= "Mature KDEL receptor inhibitor protein"  
 FT misc\_feature 97..234  
 FT /tag= d  
 FT /note= "Xenopus TSP4 trimerisation domain"  
 FT FT misc\_feature 235..306  
 FT /tag= e  
 FT /note= "Camel IgG linker domain"  
 XX  
 XX WO200006729-A1.  
 XX  
 XX 10-FEB-2000.  
 PD  
 XX  
 XX 28-JUL-1999; 99WO-US17147.  
 PF  
 XX  
 XX 29-JUL-1998; 98US-0124671.  
 PR  
 XX  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 PA  
 XX  
 XX Rothman JE, Mayhew M, Hoe MH;  
 PI WPI: 2000-195296/17.  
 DR P-PSDB; AA44958.  
 XX  
 XX Inhibitors of the KDEL receptor which comprises an oligomerization  
 PT domain useful for promoting secretion of proteins which are normally  
 PT retained within the cell -  
 PS Disclosure: Fig 5, 87pp; English.  
 XX  
 XX The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerisation domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a  
 CC cleavable signal peptide; the oligomerisation domain of Xenopus  
 CC thrombospondin 4 (TSP4) trimerisation domain including an additional  
 CC sub-sequence; a camel IgG linker domain and the carboxy-terminal sequence  
 CC KDEL. The subsequence GDCG is an alteration of rat cartilage oligomeric  
 CC matrix protein which provides increased stability via disulphide bonds.  
 XX  
 XX Sequence 369 BP; 94 A; 94 C; 121 G; 60 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1,586-25 Length: 369



Score: 371.50 Matches: 68  
 Percent Similarity: 79.41% Conservative: 13  
 Best Local Similarity: 66.67% Mismatches: 18  
 Query Match: 63.61% Indels: 3  
 DB: 21 Gaps: 1

US-09-696-872-23 (1-109) x AA250496 (1-369)

QY 8 LeuLeuAlaLeuAlaValCysSerAlaAlaLysGlySerSerLeuGlyGlyAsp 27  
 DB 40 TTGCTGCTGCTGGCGCGGTG-----CGGCGCAGAGGATCCAGCTTGAGAGAC 90  
 QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAspAlaLeu 47  
 DB 91 TGTGTGGTGCACGACAGACAGTGTGATTCGCCAGATAAACCAATGATCAGATGCTG 150  
 QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67  
 DB 151 GGAGAGCTCCAGATGTCATGACAGACAGCGTGAAAGACCAATGCTTGTAGAAACACC 210  
 QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87  
 DB 211 ATTCCAGAAATCCAGGCTGTGGCCCGCAGCCGAAACCGCACGCCGACGCCGAG 270  
 QY 88 ProGlnProLysProGlnProLysProGlnProGluProGluGlyThrGlySerSerGluLysAsp 107  
 DB 271 CCGCAGCCGAAACCGCAGCCGAAACCGAAACCGAAGGTACCGATCATCAGAAAAAGAT 330  
 QY 108 GluLeu 109  
 DB 331 GAGTTG 336

RESULT 8  
 AA250495

ID AA250495 standard; DNA; 369 BP.

XX AA250495;  
 DT 23-MAY-2000 (first entry)  
 XX KDEL receptor inhibitor-4 DNA.  
 XX KDEL receptor inhibitor; heat shock protein; immune response;  
 KM oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;  
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
 KM infectious disease; allergy; autoimmune disease; ss.  
 XX Chimeric - Mus sp.  
 OS Chimeric - Camelus sp.  
 XX Key Location/Qualifiers  
 FT CDS 10..339  
 FT /tag= a  
 FT /product= "KDEL receptor inhibitor protein"  
 FT sig\_peptide 10..68  
 FT /tag= b  
 FT /note= "Derived from mouse Bip"  
 FT mat\_peptide 69..336  
 FT /tag= c  
 FT /product= "Mature KDEL receptor inhibitor protein"  
 FT 97..234  
 FT /tag= d  
 FT /note= "Mouse TSP3 domain"  
 FT /tag= e  
 FT /note= "Camel Igg linker domain"  
 XX MO200006729-A1.  
 XX 10-FEB-2000.  
 XX 28-JUL-1999; 99WO-US171147.

PR 29-JUL-1998; 98US-0124671.  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 PA  
 XX  
 PI Rothman JE, Mayhew M, Hoe MH;  
 XX  
 DR WPI: 2000-195296/17.  
 XX P-PSDB; AAY4961.

PT Inhibitors of the KDEL receptor which comprises an oligomerization  
 PT domain useful for promoting secretion of proteins which are normally  
 PT retained within the cell  
 PS  
 XX  
 PS Disclosure; Fig 4; 87pp; English.

CC The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerisation domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a  
 CC cleavable signal peptide; the oligomerisation domain of mouse  
 CC thrombospondin 3 (TSP3) trimerisation domain including an additional  
 CC sub-sequence; a camel Igg linker domain and the carboxy-terminal sequence  
 CC KDEL. The subsequence GDC is an alteration of rat cartilage oligomeric  
 CC matrix protein which provides increased stability via disulphide bonds.

XX Sequence 369 BP; 87 A; 103 C; 120 G; 59 T; 0 other;

#### Alignment Scores:

Pred. No.: 3,71e-23 Length: 369  
 Score: 345.50 Matches: 64  
 Percent Similarity: 76.47% Conservative: 14  
 Best Local Similarity: 62.75% Mismatches: 21  
 Query Match: 59.16% Indels: 3  
 DB: 21 Gaps: 1

US-09-696-872-23 (1-109) x AA250495 (1-369)

QY 8 LeuLeuAlaLeuAlaValCysSerAlaAlaLysGlySerSerLeuGlyGlyAsp 27  
 DB 40 TTGCTGCTGCTGGCGCGGTG-----CGGCGCAGAGGATCCAGCTTGAGAGAC 90  
 QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAspAlaLeu 47  
 DB 91 TGTGTGGGAGCAGACCAAGCATGTGTCAACCGATCACTTCAACAGATCCTA 150  
 QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67  
 DB 151 GTGAGCTTCGAGACGACATCCAGACAGAGTGAAGAAATGATCATCATCCGAAACACC 210  
 QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87  
 DB 211 ATCATGAGTGTACAGTGTGGTCCGACCGCAGCCGAAACCGCAGCCGACCCGAG 270  
 QY 88 ProGlnProLysProGlnProLysProGlnProGluProGluGlyThrGlySerSerGluLysAsp 107  
 DB 271 CCGCAGCCGAAACCGCAGCCGAAACCGAAACCGAAGGTACCGATCATCAGAAAAAGAT 330  
 QY 108 GluLeu 109  
 DB 331 GAGTTG 336

RESULT 9  
 AA250494  
 ID AA250494 standard; DNA; 357 BP.

```

XX AA50494;
XX AC
XX 23-MAY-2000 (first entry)
XX DT
XX KDEL receptor inhibitor-3 DNA.
XX DE
XX KDEL receptor inhibitor; heat shock protein; immune response;
XX KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX infectious disease; allergy; autoimmune disease; ss.
XX OS
XX Chimeric - Mus sp.
XX OS
XX Chimeric - Camelus sp.
XX FT
XX Key Location/Qualifiers
XX CDS 10..327
XX FT /*tag= a
XX FT /product= "KDEL receptor inhibitor protein"
XX FT sig_peptide 10..68
XX FT /*tag= b
XX FT /note= "Derived from mouse Bip"
XX FT mat_peptide 69..324
XX FT /*tag= C
XX FT /product= "Mature KDEL receptor inhibitor protein"
XX FT misc_feature 97..222
XX FT /*tag= d
XX FT /note= "Mouse TSP3 domain"
XX FT misc_feature 223..294
XX FT /*tag= e
XX FT /note= "Camel Igg linker domain"
XX FT
XX WO200006729-A1.
XX PN
XX 10-FEB-2000.
XX PD
XX 28-JUL-1999; 99WO-US17147.
XX PF
XX 29-JUL-1998; 98US-0124671.
XX PR
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX PA
XX Rothman JE, Mayhew M, Hoe MH;
XX P1
XX WPI; 2000-195296/17.
XX DR P-PSDB; AAY44960.
XX DR
XX Inhibitor of the KDEL receptor which comprises an oligomerization
XX PT domain useful for promoting secretion of proteins which are normally
XX FT retained within the cell -
XX FT
XX PS Disclosure; Fig 3; 87pp; English.
XX XX
XX The patent discloses the use of KDEL receptor inhibitor to promote
XX CC secretion of proteins that are normally retained within the cell such as
XX CC heat shock proteins by inhibiting KDEL receptor-mediated return of
XX CC protein complexes to endoplasmic reticulum. This makes the secreted heat
XX CC shock proteins more accessible to the immune system and improves immune
XX CC response to a target antigen. The inhibitor protein comprises several
XX CC subunits where each subunit comprises an oligomerisation domain and has
XX CC at its carboxy terminus a region which binds to a KDEL receptor. The
XX CC target antigen may be associated with diseases including neoplasia such
XX CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
XX CC astrocytoma, with defective tumour suppressor genes, oncogenes,
XX CC infectious diseases, allergy or autoimmune diseases. The present
XX CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
XX CC cleavable signal peptide; the oligomerisation domain of mouse
XX CC thrombospondin 3 (TSP3) trimerisation domain; a camel Igg linker domain
XX CC and the carboxy-terminal sequence KDEL. The subsequence GPCC is an
XX CC alteration of rat cartilage oligomeric matrix protein which provides
XX CC increased stability via disulphide bonds.
XX CC
XX *Sequence 357 BP; 84 A; 100 C; 114 G; 59 T; 0 other;
XX SQ

```

```

Alignment Scores:
Pred. No.: 1.92e-22 Length: 357
Score: 337.50 Matches: 65
Percent Similarity: 75.49% Conservative: 12
Best Local Similarity: 63.73% Mismatches: 18
Query Match: 57.79% Indels: 7
DB: 21 Gaps: 2

US-09-696-872-23 (1-109) x AA50494 (1-357)
QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysblysglySerSerLeuGlyIYAAP 27
DB 40 ITGCTGCTGCTGGCGCGCTG-----CGGCCCGAGGATCCAGCTGTGGAGAC 90
QY 28 CysCysSerAspLeuGLYProGLImetLeuArgGluLeuGlnGluThrAspAlaAlaLeu 47
DB 91 TGTGTTAAGGCATTG-----GTCAACCACTCACCCTCTTACACCAATCCTTA 138
QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
DB 139 GTGGAGCTTCGGGACGACATCCGAGACCAAGTGAAGAAATGTCACTCATCCGAAACAC 198
QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87
DB 199 ATCATGAGTGTCAAGTGTGCGTCCGACCGCAGCCGAAACCGCACCCGACGCCGAG 258
QY 88 ProGlnProLysProGlnProGlnProGlnProGlnGluGlyThrGlySerSerGluLysAsp 107
DB 259 CCGGACCCGAAACCGCACCCGAAACCGAAACCGAAAGTACCGATCATCGAAAGAT 318
QY 108 GluLeu 109
DB 319 GAGTTG 324

RESULT 10
AAZ50498
ID AAZ50498 standard; DNA; 312 BP.
XX AC
XX AAZ50498;
XX DT 23-MAY-2000 (first entry)
XX DE
XX KDEL receptor inhibitor-7 DNA.
XX KW KDEL receptor inhibitor; heat shock protein; immune response;
XX KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
XX KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX KW infectious disease; allergy; autoimmune disease; ss.
XX OS
XX Chimeric - Adenovirus E3.
XX OS
XX Chimeric - Homo sapiens.
XX OS
XX Chimeric - Camelus sp.
XX FT
XX Key Location/Qualifiers
XX FT sig_peptide 10..69
XX FT /*tag= a
XX FT /note= "Derived from adenovirus E3"
XX FT CDS 10..282
XX FT /*tag= b
XX FT /note= "KDEL receptor inhibitor protein"
XX FT mat_peptide 70..279
XX FT /*tag= c
XX FT /product= "Mature KDEL receptor inhibitor protein"
XX FT misc_feature 97..177
XX FT /*tag= d
XX FT /note= "Human PLB pentamerisation domain"
XX FT misc_feature 178..249
XX FT /*tag= e
XX FT /note= "Camel Igg linker domain"
XX PN
XX WO200006729-A1.

```





XX Diagnosing breast cancer in a patient comprises detecting the level of  
PT gene expression in cell or tissue samples, where a differential gene  
PT expression is indicative of breast cancer -  
PS Claim 1; SEQ ID NO 1021; 260pp + Sequence Listing; English.  
XX The present invention relates to methods of diagnosing breast cancer in a  
CC patient, which comprise detecting the level of expression in a tissue  
CC sample of two or more genes selected from those shown in ABT09867-  
CC ABT1112, where a differential expression of the genes indicates breast  
CC cancer. The methods are useful in diagnosing, treating, detecting the  
CC progression, and in monitoring treatment of breast cancer in patients.  
CC The methods are also useful as a screening tool for agents that modulate  
CC the onset or progression of breast cancer. The breast cancer genes may be  
CC used as diagnostic markers for the prediction or identification of the  
CC malignant state of breast tissue, for confirming the type and progression  
CC of cancer, and for drug screening and assays. The present sequence is a  
CC coding sequence of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2439 BP; 503 A; 758 C; 809 G; 369 T; 0 other;

Alignment Scores:

Pred. No.: 1.14e-14 Length: 2439  
Score: 262.50 Matches: 58  
Percent Similarity: 76.47% Conservative: 7  
Best Local Similarity: 68.24% Mismatches: 16  
Query Match: 44.95% Indels: 5  
DB: 24 Gaps: 1

US-09-696-872-23 (1-109) x ABT10887 (1-2439)

QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaIleValysGlySerSerLeuGlyAsp 27  
DB 53 CTGCTCACCCCTGGCTGCGCTCGCGCCGCGGACAGGCGCCAGAGCCGCTTGGGC----- 106  
QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgIleuGlnGluThrAsnAlaIleu 47  
DB 107 -----TCAGACCTGGGCGCCGACAGTGTCTCGGAACTGCAGAAACCAACGCGCGCTG 160  
QY 48 GlnAspValArgAspTPrLeuArgGlnGlnValArgGluIleThrPheLeuIleAsnThr 67  
DB 161 CAGGACGTGGGAGCTGCTCGGACAGGTCAGGGAGATCAGTCTTGAATAAACG 220  
QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87  
DB 221 GTGATGAGTGTGACGCGCTGCGGATGACAGCAGTCAATACGAC-CGCGCTACCCAGCGT 279

QY 88 ProGlnProIysPro 92  
DB 280 GCGGCCCTGCTCCA 294

RESULT 14  
ABT07751  
ID ABT07751 standard; DNA; 2439 BP.

XX ABT07751;  
XX  
XX 14-NOV-2002 (first entry)

DE Breast cancer-associated gene sequence 59.

KM Gene; ds; breast cancer; breast cancer-associated gene sequence;  
KM drug development; pharmacogenetics; biosensor development.

XX Unidentified.

XX WO200259377-A2.

XX 01-AUG-2002.

XX 24-JAN-2002; 2002WO-US02242.  
XX  
XX 24-JAN-2001; 2001US-263965P.  
XX 02-FEB-2001; 2001US-265928P.  
XX 09-APR-2001; 2001US-282698P.  
XX 09-APR-2001; 2001US-0829472.  
XX 04-MAY-2001; 2001US-288590P.  
XX 29-MAY-2001; 2001US-294443P.  
XX (EOSB-) EOS BIOTECHNOLOGY INC.  
XX  
XX Mack DH, Gish KC, Afar D;  
XX WPI: 2002-583738/62.  
XX N-PSDB; ABU05594.  
XX  
XX Detecting a breast cancer-associated transcript in a patient's cell,  
XX useful for diagnosing breast cancer, comprises contacting a biological  
XX sample with a polynucleotide that selectively hybridizes with breast  
XX cancer nucleic acids -  
XX  
XX Claim 9; Page 400-401; 410pp; English.

XX The invention comprises a method of detecting a breast cancer-associated  
XX transcript in a cell from a patient. The method of the invention involves  
XX contacting a biological sample from the patient with a nucleotide that  
XX hybridizes to one of the 69 breast cancer-associated gene sequences shown  
XX in the specification. The method of the invention is useful in the  
XX diagnosis or prognosis of breast cancer, and for detecting genes that are  
XX up or down-regulated in breast cancer cells. Genes identified by the  
XX method of the invention can be used in diagnostic purposes and also as  
XX targets for screening for therapeutic compounds that modulate breast  
XX cancer (e.g. hormones or antibodies). Identification of genes that are  
XX over or under expressed in breast cancer can additionally provide high-  
XX resolution, high-sensitivity datasets which can be used in the areas of  
XX diagnostics, therapeutics, drug development, pharmacogenetics, protein  
XX structure and biosensor development. DNA sequences ABT07693 - ABT07761  
XX represent the 69 breast cancer-associated gene sequences of the  
XX invention.

XX Sequence 2439 BP; 503 A; 758 C; 809 G; 369 T; 0 other;

Alignment Scores:

Pred. No.: 1.14e-14 Length: 2439  
Score: 262.50 Matches: 58  
Percent Similarity: 76.47% Conservative: 7  
Best Local Similarity: 68.24% Mismatches: 16  
Query Match: 44.95% Indels: 5  
DB: 24 Gaps: 1

US-09-696-872-23 (1-109) x ABT07751 (1-2439)

QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaIleValysGlySerSerLeuGlyAsp 27  
DB 53 CTGCTCACCCCTGGCTGCGCTCGCGCCGCGGACAGGCGCCAGAGCCGCTTGGGC----- 106  
QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgIleuGlnGluThrAsnAlaIleu 47  
DB 107 -----TCAGACCTGGGCGCCGACAGTGTCTCGGAACTGCAGAAACCAACGCGCGCTG 160  
QY 48 GlnAspValArgAspTPrLeuArgGlnGlnValArgGluIleThrPheLeuIleAsnThr 67  
DB 161 CAGGACGTGGGAGCTGCTCGGACAGGTCAGGGAGATCAGTCTTGAATAAACG 220  
QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87  
DB 221 GTGATGAGTGTGACGCGCTGCGGATGACAGCAGTCAATACGAC-CGCGCTACCCAGCGT 279

QY 88 ProGlnProIysPro 92

DB 280 GCGGCCCTGCTCCA 294

RESULT 15
ABL62104
ID ABL62104 standard; DNA, 2439 BP.
XX AC ABL62104;
XX DT 15-MAY-2002 (first entry)
DE Colon adenocarcinoma related gene sequence SEQ ID NO:441.
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; carcinosarcoma;
KM cytoskeletal; gene therapy; anti-neoplastic; Wilms' tumour; adenocarcinoma;
gene; ds.
XX Homo sapiens.
OS
XN WO200194629-A2.
PD 13-DEC-2001.
PF 30-MAY-2001; 2001WO-US10838.
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-231131P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-231172P.
PR 02-OCT-2000; 2000US-231173P.
PR 02-OCT-2000; 2000US-231278P.
PR 02-OCT-2000; 2000US-231294P.
PR 02-OCT-2000; 2000US-231295P.
PR 02-OCT-2000; 2000US-231316P.
PR 03-OCT-2000; 2000US-231425P.
PR 03-OCT-2000; 2000US-231598P.
PR 03-OCT-2000; 2000US-231604P.
PR 03-OCT-2000; 2000US-231606P.
PR 03-OCT-2000; 2000US-231608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX PA (AVALON PHARM.
XX PI Young PE, Augustine M, Carter KC, Ebner R, Endress G, Horrigan S,
EI Soppe DR, Weaver Z;
XX .MPI; 2002-188264/24.
XX

PT Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
XX  
XX  
PS Claim 1, SEQ ID 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000, 1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012, 1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032, 1033, 1034, 1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045, 1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056, 1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067, 1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078, 1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089, 1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100, 1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111, 1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122, 1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133, 1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144, 1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155, 1156, 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166, 1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177, 1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188, 1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199, 1200, 1201, 1202, 1203, 1204, 1205,

Alignment Scores:	
Pred. No.:	1,14e-14
Score:	262.50
Percent Similarity:	76.47%
Best Local Similarity:	68.24%
Query Match:	44.93%
DB:	24
US-09-696-872-23 (1-109) x ABL62104 (1-2439)	
	Length: 2439
	Matches: 58
	Conservative: 7
	Mismatches: 16
	Indels: 5
	Gaps: 1

[illegible]

Search completed: August 26, 2003, 17:06:37  
Job time : 376.301 secs

PA (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S,  
PI Soppet DR, Weaver Z;  
XX  
XX  
DR ,MPI; 2002-188264/24.  
XX

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 26, 2003, 16:55:35 ; Search time 85.8496 Seconds

(without alignments)  
560.408 Million cell updates/sec

Title: US-09-696-872-23  
Perfect score: 584  
Sequence: 1 MRVWLLGLALAAVCSAAKK.....PKPQKPEPEGTGSSEKDEL 109

Scoring table: BLOSUM62  
Xgapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09696872/runat\_26082003\_151138\_3255/app\_query.fasta\_1.462  
-DB=issued.patents.NA -QEMT=fastap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=PCT -NOM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=2000000000  
-USER=US09696872.QCN 1.1 76.@runat\_26082003\_151138\_3255 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEDBTRY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRAD=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfltest.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	579	99.1	372	3	US-09-124-671-24
2	459	78.6	372	3	US-09-124-671-30
3	429	73.5	372	3	US-09-124-671-28
4	428.5	73.4	387	3	US-09-124-671-16
5	422	72.3	444	3	US-09-124-671-35
6	414.5	71.0	387	3	US-09-124-671-14
7	371.5	63.6	369	3	US-09-124-671-22
8	345.5	59.2	357	3	US-09-124-671-20
9	337.5	57.8	357	3	US-09-124-671-18
10	304.5	52.1	315	3	US-09-124-671-26
11	145.5	24.9	3074	5	PCT-US93-11725-3
12	138.5	23.7	2820	5	PCT-US93-11725-1

13	122	20.9	521	4	US-09-643-597-324	Sequence 324, App
14	122	20.9	521	4	US-09-480-884A-324	Sequence 324, App
15	122	20.9	521	4	US-09-542-615A-324	Sequence 324, App
16	122	20.9	521	4	US-09-606-421B-324	Sequence 324, App
17	119	20.4	2790	2	US-08-718-661-1	Sequence 1, Appl1
18	118	20.2	17136	4	US-09-453-702B-158	Sequence 158, App
19	109	18.7	816	4	US-09-252-991A-721	Sequence 721, App
20	109	18.7	1605	4	US-09-252-991A-694	Sequence 694, App
21	109	18.7	1674	4	US-09-252-991A-689	Sequence 689, App
22	109	18.7	1677	4	US-09-252-991A-631	Sequence 631, App
23	108.5	18.6	2629	4	US-09-332-184-17	Sequence 17, Appl
24	108	18.5	405	4	US-09-252-991A-6532	Sequence 6532, App
25	108	18.5	546	4	US-09-252-991A-6514	Sequence 6514, App
26	108	18.5	591	4	US-09-252-991A-6545	Sequence 6545, App
27	108	18.5	915	4	US-09-252-991A-6545	Sequence 6545, App
28	108	18.5	2765	4	US-09-620-112D-61	Sequence 61, Appl
29	107	18.3	9758	4	US-09-332-812A-5	Sequence 5, Appl1
30	103	17.6	10348	2	US-08-457-273B-41	Sequence 41, Appl
31	103	17.6	10348	3	US-08-556-419-13	Sequence 13, Appl
32	103	17.6	10348	3	US-09-041-886-14	Sequence 14, Appl
33	103	17.6	10366	1	US-08-246-982A-5	Sequence 5, Appl1
34	103	17.6	10366	1	US-08-453-265-5	Sequence 5, Appl1
35	101	17.3	450	4	US-09-252-991A-2601	Sequence 2601, App
36	101	17.3	588	4	US-09-252-991A-2243	Sequence 2243, App
37	101	17.3	732	4	US-09-252-991A-2440	Sequence 2440, App
38	101	17.3	1083	4	US-09-252-991A-2517	Sequence 2517, App
39	101	17.3	1431	4	US-09-252-991A-2397	Sequence 2397, App
40	100.5	17.2	2369	2	US-08-525-742-1	Sequence 1, Appl1
41	100	17.1	9758	4	US-09-332-812A-5	Sequence 5, Appl1
42	99	17.0	1144	1	US-08-014-943A-1	Sequence 1, Appl1
43	99	17.0	1144	1	US-08-486-421-2	Sequence 2, Appl1
44	99	17.0	1144	1	US-08-470-911-2	Sequence 2, Appl1
45	99	17.0	1144	2	US-08-486-809-2	Sequence 2, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-124-671-24  
Sequence 24, Application US/09124671A  
Patent No. 6160088

GENERAL INFORMATION:  
APPLICANT: Rothman, James  
APPLICANT: Mayhew, Mark  
TITLE OR INVENTION: KDEL RECEPTOR INHIBITORS  
FILE REFERENCE: 31488  
CURRENT FILING DATE: 1998-07-29  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatsSeq for Windows Version 3.0  
SEQ ID NO 24  
LENGTH: 372  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chimeric human COMP-KDEL  
US-09-124-671-24

Alignment Scores:  
Pred. No.: 3.03e-56  
Score: 579.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 99.14%  
DB: 3  
Gaps: 0  
Length: 372  
Matches: 108  
Conservative: 0  
Mismatch: 0  
Indels: 0

US-09-696-872-23 (1-109) x US-09-124-671-24 (1-372)

QY 2 ArgTyrMetIleLeuGlyLeuAlaLeuAlaValCysSerAlaAlaIleValGlyGly 21  
DB 16 AGGTAATGATTTTAAAGCTGCTGCGCCCTTGCGGCACTGCGAGCGCTGCCAAAAAGGA 75

```
QY 22 SerSerLeuGlyGlyAspCySerAspLeuGlyProGlnMetLeuArgGluLeuGln 41
|
|
|
Db 76 TCACCTCGGTGGAGACTGTGTCAGACCTGGGCCCGCAGACTGCTTGGGAACTGCAG 135
|
|
|
QY 42 GluThrAsnAlaAlaLeuGlnAspValArgAspTrpLeuArgGlnValArgGluLe 61
|
|
|
Db 136 GAAACCAAGCGCGCTGAGAGACGGCGGAGCTGCTCGGCAGAGGTCAAGGAGATC 195
|
|
|
QY 62 ThrPheLeuLysAsnThrValMetGluCyAspAlaCyGlyProGlnProGlnProLys 81
|
|
|
Db 196 ACGTTCCTGAAAAACACGGGTGATGAGTGTGACGGCTGGGCCCGCAGCCGAGCCGAAA 255
|
|
|
QY 82 ProGlnProGlnProGlnProGlnProLysProGlnProLysProGlnProGluGlyThr 101
|
|
|
Db 256 CCGGAGCCGCGACCCCGACGCCGAGCCGAAACCGCAGCCGAAACCGGAACCGGAAGGTACC 315
|
|
|
QY 102 GlySerSerGluLysAspGluLeu 109
|
|
|
Db 316 GGATCATCAGAAAAAGATGAGTTG 339
|
|
|
RESULT 2
US-09-124-671-30
; Sequence 30, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric human TSP4-KDEL
US-09-124-671-30
Alignment Scores:
Pred. No.: 6,96e-43 Length: 372
Score: 459.00 Matches: 84
Percent Similarity: 86.11% Conservative: 9
Best Local Similarity: 77.78% Mismatches: 15
Query Match: 78.60% Indels: 0
Gaps: 0
DB: 3
US-09-696-872-23 (1-109) x US-09-124-671-30 (1-372)
QY 2 ArgTyrMetIleLeuGlyLeuLeuAlaLeuAlaValCySerAlaAlaLysGly 21
|
|
|
Db 16 AGGTACATGATTTTAAAGCTTCTCGCCCTTGGGAGGTGAGCGCTGCCAAAAAGGA 75
|
|
|
QY 22 SerSerLeuGlyGlyAspCySerAspLeuGlyProGlnMetLeuArgGluLeuGln 41
|
|
|
Db 76 TCACCTCGGTGGAGACTGTGTCAGACCTTAAACCGGCACTTCTGGGTCAAAATGACA 135
|
|
|
QY 42 GluThrAsnAlaAlaLeuGlnAspValArgAspTrpLeuArgGlnValArgGluLe 61
|
|
|
Db 136 CAATTAAACCAACTCTGGAGAGGTGAAAGACCTTCTGAGACACAGCTTAAGGAAACA 195
|
|
|
QY 62 ThrPheLeuLysAsnThrValMetGluCyAspAlaCyGlyProGlnProGlnProLys 81
|
|
|
Db 196 TCATTTTTCGAAAAACCATACCTGTAATGCCAGGCTTGGGTCGCCGAGCCGAGCGAAA 255
|
|
|
QY 82 ProGlnProGlnProGlnProGlnProLysProGlnProLysProGlnProGluGlyThr 101
|
|
|
Db 256 CCGGAGCCGCGACCCCGACGCCGAGCCGAAACCGCAGCCGAAACCGGAACCGGAAGGTACC 315
|
|
|
QY 102 GlySerSerGluLysAspGluLeu 109
|
|
|
Db 316 GGATCATCAGAAAAAGATGAGTTG 109
```

```
Db 316 GGATCATCAGAAAAAGATGAGTTG 339
|
|
|
RESULT 3
US-09-124-671-28
; Sequence 28, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric human TSP3-KDEL
US-09-124-671-28
Alignment Scores:
Pred. No.: 1.52e-39 Length: 372
Score: 429.00 Matches: 78
Percent Similarity: 84.26% Conservative: 13
Best Local Similarity: 72.22% Mismatches: 17
Query Match: 73.46% Indels: 0
Gaps: 0
DB: 3
US-09-696-872-23 (1-109) x US-09-124-671-28 (1-372)
QY 2 ArgTyrMetIleLeuGlyLeuLeuAlaLeuAlaValCySerAlaAlaLysGly 21
|
|
|
Db 16 AGGTACATGATTTTAAAGCTTCTCGCCCTTGGGAGGTGAGCGCTGCCAAAAAGGA 75
|
|
|
QY 22 SerSerLeuGlyGlyAspCySerAspLeuGlyProGlnMetLeuArgGluLeuGln 41
|
|
|
Db 76 TCACCTCGGTGGAGACTGTGTCAGACCTTAAACCGGCACTTCTGGGTCAAAATGACA 135
|
|
|
QY 42 GluThrAsnAlaAlaLeuGlnAspValArgAspTrpLeuArgGlnValArgGluLe 61
|
|
|
Db 136 CTCTTCAACCAAGATCCTGAGAGCTTGGGAGCGACATCCGAGACCAAGGTGAAGGAATG 195
|
|
|
QY 62 ThrPheLeuLysAsnThrValMetGluCyAspAlaCyGlyProGlnProGlnProLys 81
|
|
|
Db 196 TCATCATCCGGAACCATCATGAGTGTGAGGTGTCGGTCCGACGCCGAGCCGAAA 255
|
|
|
QY 82 ProGlnProGlnProGlnProGlnProLysProGlnProLysProGlnProGluGlyThr 101
|
|
|
Db 256 CCGGAGCCGCGACCCCGACGCCGAGCCGAAACCGCAGCCGAAACCGGAACCGGAAGGTACC 315
|
|
|
QY 102 GlySerSerGluLysAspGluLeu 109
|
|
|
Db 316 GGATCATCAGAAAAAGATGAGTTG 339
|
|
|
RESULT 4
US-09-124-671-16
; Sequence 16, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
```



SEQ ID NO 16  
 LENGTH: 387  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: chimeric rat COMP-KDEL  
 US-09-124-671-16

## Alignment Scores:

Pred. No.: 1.82e-39 Length: 387  
 Score: 428.50 Matches: 87  
 Percent Similarity: 80.36% Conservative: 3  
 Best Local Similarity: 77.68% Mismatches: 5  
 Query Match: 73.37% Indels: 17  
 DB: 3 Gaps: 3

US-09-696-872-23 (1-109) x US-09-124-671-16 (1-387)

Qy 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaValSerGlySerSerLeuGlyValAsp 27  
 Db 40 TTGCTGCTGCTGGGGCGGCTG-----CGGGCCAGAGGATCCAGCTTGGTGGAGAC 90  
 Qy 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaLeu 47  
 Db 91 TGTGT-----CCACAGATGCTTCAGAGAACTCCAGAGACTAATGCGCGCTG 138  
 Qy 48 GlAspValArgAspTrpLeuArgGlnValArgGluIleThrPheLeuValAsnThr 67  
 Db 139 CAAGACGTGAGAGAGCTCTTCGACAGCAGGTCAAGAGATACCTTCTGAAGAAATCG 198  
 Qy 68 ValMetGluCysAspAlaCysGly-----ProGln 77  
 Db 199 GTGATGAAATGTGACGCTTGCGAATGCAAGCCGACGCCGCGGATCAATGCCGAG 258  
 Qy 78 ProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProLysProGlu 97  
 Db 259 CCGCAGCCGAAACCGCAGCCGACCGCAGCCGCGAAGCGCAACCGCAACCGGAA 318  
 Qy 98 ProGluGlyThrGlySerSerGluLysAspGluLeu 109  
 Db 319 CCGGAGGTACCGGATCATCAAGAAAGATGATTG 354

## RESULT 5

US-09-124-671-35  
 Sequence 35, Application US/09124671A

Patent No. 6160088  
 GENERAL INFORMATION:  
 APPLICANT: Rothman, James  
 APPLICANT: Mayhew, Mark  
 APPLICANT: Hoe, Mee  
 TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS  
 FILE REFERENCE: 31488  
 CURRENT APPLICATION NUMBER: US/09/124,671A  
 CURRENT FILING DATE: 1998-07-29  
 NUMBER OF SEQ ID NOS: 42  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 35  
 LENGTH: 444  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: KDEL-myc  
 US-09-124-671-35

## Alignment Scores:

Pred. No.: 1.15e-38 Length: 444  
 Score: 422.00 Matches: 82  
 Percent Similarity: 84.85% Conservative: 2  
 Best Local Similarity: 82.83% Mismatches: 1  
 Query Match: 72.26% Indels: 14  
 DB: 3 Gaps: 2

US-09-696-872-23 (1-109) x US-09-124-671-35 (1-444)

Qy 21 GlySerSerLeuGlyGlyAspCysCysSerAspLeuGlyProGlnMetLeuArgGluLeu 40  
 Db 127 GATCCAGCTTGGGTGAGACTGTGT-----CCACAGATGCTTCAGAACTC 174  
 Qy 41 GlnGluThrAsnAlaAlaLeuGlnAspValArgAspTrpLeuArgGlnValArgGlu 60  
 Db 175 CAGGAGACTAATGCGGCGCTGCAAGACGTGAGAGAGCTTGGCAGCAGGTCAAGAG 234  
 Qy 61 IleThrPheLeuLysAsnThrValMetGluCysAspAlaCysGly----- 75  
 Db 235 ATCACTTCTTGAAGAAATGAGTGAATGATGACGCTTCGGATGACGCCGACGC 294  
 Qy 76 -----ProGlnProGlnProLysProGlnProGlnProGlnPro 90  
 Db 295 ACCCGCGTACTAGTCCGACGCCGCGACCCGAAACCGACCGCAGCCGCGACCGC 354  
 Qy 91 LysProGlnProLysProGlnProGlnGlyThrGlySerSerGluLysAspGluLeu 109  
 Db 355 AAACCGCAGCGCAACCGGAACCGGAAGTACCGGATCATCAAGAAAGATGAGTTG 411

## RESULT 6

US-09-124-671-14  
 Sequence 14, Application US/09124671A

Patent No. 6160088  
 GENERAL INFORMATION:  
 APPLICANT: Rothman, James  
 APPLICANT: Mayhew, Mark  
 APPLICANT: Hoe, Mee  
 TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS  
 FILE REFERENCE: 31488  
 CURRENT APPLICATION NUMBER: US/09/124,671A  
 CURRENT FILING DATE: 1998-07-29  
 NUMBER OF SEQ ID NOS: 42  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 14  
 LENGTH: 387  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: chimeric rat COMP-KDEL  
 US-09-124-671-14

## Alignment Scores:

Pred. No.: 6.61e-38 Length: 387  
 Score: 414.50 Matches: 86  
 Percent Similarity: 79.46% Conservative: 3  
 Best Local Similarity: 76.79% Mismatches: 6  
 Query Match: 70.98% Indels: 17  
 DB: 3 Gaps: 3

US-09-696-872-23 (1-109) x US-09-124-671-14 (1-387)

Qy 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaValSerGlySerSerLeuGlyValAsp 27  
 Db 40 TTGCTGCTGCTGGGGCGGCTG-----CGGGCCAGAGGATCCAGCTTGGTGGAG 87  
 Qy 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaLeu 47  
 Db 88 -----GACCTAGCCCAACAGATGCTTCGAGAACTCCAGAGACTAATGCGCGCTG 138  
 Qy 48 GlAspValArgAspTrpLeuArgGlnValArgGluIleThrPheLeuValAsnThr 67  
 Db 139 CAAGACGTGAGAGAGCTCTTCGACAGCAGGTCAAGAGATCACTTCTGAAGAAATCG 198  
 Qy 68 ValMetGluCysAspAlaCysGly-----ProGln 77  
 Db 199 GTGATGAAATGTGACGCTTGCGAATGCAAGCCGACGCCGCGGATCAATGCCGAG 258  
 Qy 78 ProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProLysProGlu 97  
 Db 259 CCGCAGCCGAAACCGCAGCCGACCGCAGCCGCGAAGCGCAACCGCAACCGGAA 318

```
OY 98 ProGluGlyThrGlySerSerGluLysAspGluLeu 109
DB 319 CCGAAGGTACCGGATCATCAGAAAAAGATGAGTTG 354

RESULT 7
US-09-124-671-22
; Sequence 22, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 369
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric Xenopus laevis TSP4-KDEL
US-09-124-671-22

Alignment Scores:
Pred. No.: 3,81e-33 Length: 369
Score: 371.50 Matches: 68
Percent Similarity: 79.41% Conservative: 13
Best Local Similarity: 66.67% Mismatches: 18
Query Match: 63.61% Indels: 3
DB: 3 Gaps: 1

US-09-696-872-23 (1-109) x US-09-124-671-22 (1-369)
OY 8 LeuLeuAlaLeuAlaAlaValCySerAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27
DB 40 TTGCTGCTGCTGGGCGCGGTG-----CGGGCCAGGGATCCAGCTGGGTGAGAGC 90
OY 28 CyCySerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
DB 91 TGTGTGTGTGAGTCAAGACGAGATTGATGGCCAGATTAACCAATGATGATGCTG 150
OY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluLeuThrPheLeuLysAsnThr 67
DB 151 GGAGAGCTCCGAGATGTCTATGAGACAGCGTGAAAGAGACCATGTTCTTGAGAAACACC 210
OY 68 ValMetGluCyAspAlaCyGlyProGlnProGlnProGlnProGlnProGln 87
DB 211 ATTGACAGATGCCAGGCTGTGGCCCGGACCGGACCGGAAACCCGACGCCGACG 270
OY 88 ProGlnProLysProGlnProLysProGlnProGluGlyThrGlySerSerGluLysAsp 107
DB 271 CCGGACCGGAAACCCGACCGGAAACCGGAAACCGGAAAGTACCGGATCATCAGAAAAAGAT 330
OY 108 GluLeu 109
DB 331 GAGTTG 336

RESULT 8
US-09-124-671-20
; Sequence 20, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 20
; LENGTH: 369
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric mouse TSP3-KDEL
US-09-124-671-20

Alignment Scores:
Pred. No.: 2.99e-30 Length: 369
Score: 345.50 Matches: 64
Percent Similarity: 76.47% Conservative: 14
Best Local Similarity: 62.75% Mismatches: 21
Query Match: 59.16% Indels: 3
DB: 3 Gaps: 1

US-09-696-872-23 (1-109) x US-09-124-671-20 (1-369)
OY 8 LeuLeuAlaLeuAlaAlaValCySerAlaAlaLysLysGlySerSerLeuGlyGlyAsp 27
DB 40 TTGCTGCTGCTGGGCGCGGTG-----CGGGCCAGGGATCCAGCTGGGTGAGAGC 90
OY 28 CyCySerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
DB 91 TGTGTGTGTGAGTCAAGACGAGATTGATGGCCAGATTAACCAATGATGATGCTG 150
OY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluLeuThrPheLeuLysAsnThr 67
DB 151 GTGAGGCTTCGGAGACGACATCCGAGACCGAGTGAAAGAAATGATCATCTCAGGAACACC 210
OY 68 ValMetGluCyAspAlaCyGlyProGlnProGlnProGlnProGlnProGln 87
DB 211 ATCATGAGTGTCAAGTGTGTGGTCCGACGCCGACCGGAAACCCGACGCCGACG 270
OY 88 ProGlnProLysProGlnProLysProGlnProGluGlyThrGlySerSerGluLysAsp 107
DB 271 CCGGACCGGAAACCCGACCGGAAACCGGAAACCGGAAAGTACCGGATCATCAGAAAAAGAT 330
OY 108 GluLeu 109
DB 331 GAGTTG 336

RESULT 9
US-09-124-671-18
; Sequence 18, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 357
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric mouse TSP3-KDEL
US-09-124-671-18

Alignment Scores:
Pred. No.: 2.23e-29 Length: 357
Score: 337.50 Matches: 65
Percent Similarity: 75.49% Conservative: 12
Best Local Similarity: 63.73% Mismatches: 18
Query Match: 57.79% Indels: 7
DB: 3 Gaps: 2
```

```

0y      8/ GlnPtoGlnPtoLysPto 96
      ||| |||
050 GGGGGG---GGGGGGGGGGGGGGGG 005

```

RESULT 12  
 PCT-US93-11725-1  
 ; Sequence 1, Application PC/TUS9311725  
 ; GENERAL INFORMATION:  
 ; APPLICANT:  
 ; TITLE OF INVENTION: HUMAN THROMBOSPONDIN-4  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Wolf, Greenfield, & Sacks, P.C.  
 ; STREET: 600 Atlantic Avenue  
 ; CITY: Boston  
 ; STATE: Massachusetts  
 ; COUNTRY: United States of America  
 ; ZIP: 02210  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3 1/2 inch  
 ; COMPUTER: IBM-compatible  
 ; OPERATING SYSTEM: MS-DOS Version 3.3  
 ; SOFTWARE: WordPerfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US93/11725  
 ; FILING DATE: filed herewith  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 07/985,296  
 ; FILING DATE: 04-DEC-1992  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: GATES, Edward R.  
 ; REGISTRATION NUMBER: 31,616  
 ; REFERENCE/DOCKET NUMBER: B0801/7005WO  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2820 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; ANTI-SENSE: no  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Xenopus laevis  
 ; DEVELOPMENTAL STAGE: Stage 45 (germ line)  
 ; PCT-US93-11725-1  
  
 Alignment Scores:  
 Pred. No.: 4.63e-06 Length: 2820  
 Score: 138.50 Matches: 37  
 Percent Similarity: 46.30% Conservative: 13  
 Best Local Similarity: 34.26% Mismatches: 31  
 Query Match: 23.72% Indels: 27  
 DB: 5 Gaps: 4  
  
 US-09-696-872-23 (1-109) x PCT-US93-11725-1 (1-2820)  
 QY 23 Serleuuglyglaasgcyserseerpleuuglyproglinmetleuargluleuuglu 42  
 DB 424 AGTGAAGCCGCGACAGACAGTGTACGTCACAGACAGATTGATGGCCAGTAACCA 483  
 QY 43 Thraenalaaleuuglnasvalargaprtpleuarglnlnvalargsluierthr 62  
 DB 484 ATGATTCAGATCTCTGGAGAGCTCCGAGATCTCAGACAGACAGCTGAAAGACCATG 543  
 QY 63 Pheleuysaerthvalmetglucysapalacygylproglinproglin---Prolys 81  
 DB 544 TTCTTGAGAAACACCATTCAGAAATGCCAGGCTGTGCTTGAATCTTCCATTG 603  
 QY 82 Proglinproglinproglin-----Proglinprolypro----- 92  
 DB 604 CCAACCAAGTTCCCGACGCGCTACCACTACACCTCCCAAGCTTCGATGATGCA 663  
 QY 93 -----glnprolypro-----glnprolypro 96  
 DB 664 ACTTATGTTTCAGAGAGTGGGTCATTGATACAGAGGGGGCTTCCATGTTGGCCG 723

QY 97 GluProglinly---Thnglyser 103  
 DB 724 TGTCCTGAAGCTATACAGGCAAC 747  
  
 RESULT 13  
 US-09-643-597-324  
 ; Sequence 324, Application US/09643597  
 ; Patent No. 6426072  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Tonglong  
 ; APPLICANT: Fan, Liqun  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Bangur, Chaitanya S.  
 ; APPLICANT: Hoeken, Nancy  
 ; APPLICANT: Fanger, Gary R.  
 ; APPLICANT: Li, Samuel X.  
 ; APPLICANT: Wang, Aijun  
 ; APPLICANT: Skeiky, Yasir A.W.  
 ; APPLICANT: Henderson, Robert A.  
 ; APPLICANT: McNeill, Patricia D.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; FILE REFERENCE: 210121.455C11  
 ; CURRENT APPLICATION NUMBER: US/09/643,597  
 ; CURRENT FILING DATE: 2000-08-21  
 ; NUMBER OF SEQ ID NOS: 369  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 324  
 ; LENGTH: 521  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; US-09-643-597-324

Alignment Scores:  
 Pred. No.: 3.59e-05 Length: 521  
 Score: 122.00 Matches: 19  
 Percent Similarity: 95.83% Conservative: 4  
 Best Local Similarity: 79.17% Mismatches: 1  
 Query Match: 20.89% Indels: 0  
 DB: 4 Gaps: 0

US-09-696-872-23 (1-109) x US-09-643-597-324 (1-521)

QY 76 Proglinproglinprolyproglinproglinproglinprolyproglinprolys 95  
 DB 189 CCTCAGCTCAGCCGCAACCCCAATCACACCCCAAGCTCAGCCCAACCCCAAG 248  
 QY 96 Proglinproglin 99  
 DB 249 CCTCAGCCCGCAG 260  
  
 RESULT 14  
 US-09-480-884A-324  
 ; Sequence 324, Application US/09480884A  
 ; Patent No. 6482597  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Tonglong  
 ; APPLICANT: Fan, Liqun  
 ; APPLICANT: Hoeken, Nancy A.  
 ; APPLICANT: Kalos, Michael D.  
 ; APPLICANT: Fanger, Gary R.  
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY  
 ; FILE REFERENCE: 210121.455C6  
 ; CURRENT APPLICATION NUMBER: US/09/480,884A  
 ; CURRENT FILING DATE: 2001-08-27  
 ; NUMBER OF SEQ ID NOS: 330  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 324  
 ; LENGTH: 521  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; US-09-480-884A-324

Alignment Scores:

Pred. No.:	3.59e-05	length:	521
Score:	122.00	Matches:	19
Percent Similarity:	95.83%	Conservative:	4
Best Local Similarity:	79.17%	Mismatches:	1
Query Match:	20.89%	Indels:	0
DB:	4	Gaps:	0

US-09-696-872-23 (1-109) X US-09-480-884A-324 (1-521)

**QY** 76 ProGlnProGlnProLysProGlnProGlnProGlnProGlnProLys 95  
|||||  
||::|||  
**Db** 189 CCTCAGCTCAGCCGCACCCAGCCCCAATCAACAACCCAGCTTCAGCCCCAACCCAAAG 248  
|||||

Qy	96	ProGluProGlu	99
		::   :::	
Db	249	CCTCAGCCCCAG	260

RESULT 15  
HS-09-543

US-09-542-615A-324  
Sequence 324, Application US/09542615A  
Patent No. 6518256  
GENERAL INFORMATION:  
APPLICANT: Wang, Tongtong  
APPLICANT: Fan, Liqun  
APPLICANT: Kalos, Michael D.  
APPLICANT: Bangur, Chantanya S.  
APPLICANT: Hosken, Nancy A.  
APPLICANT: Faager, Gary R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
FILE REFERENCE: 210121, 45568  
CURRENT APPLICATION NUMBER: US/09/542,615A  
CURRENT FILING DATE: 2000-04-14  
NUMBER OF SEQ ID NOS: 350  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 324  
LENGTH: 521  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-542-615A-324

Alignment Scores:

Pred. No.:	3,596-05	length:	528
Score:	122.00	Matches:	19
Percent Similarity:	95.83%	Conservative:	4
Best Local Similarity:	79.17%	Mismatches:	1
Query Match:	20.89%	Indels:	0
DB:	4	Gaps:	0

US-09-696-872-23 (1-109) x US-09-542-615A-324 (1-521).

```
QY      76 ProGlnProGlnProLysProGlnProGlnProGlnProLys 95
       |||||:::|||||:::|||||:::|||||
Db     189 CCTCAGCTCAGCCGCACCCAGCCCAATCACAAACCCAGCTTCAGCCCAACCCAAAG 248
```

Qy	96	ProGluProGlu	99
		::   :::	
Db	249	CCTCAGCCCCAG	260

Search completed: August 26, 2003, 19:22:49  
Job time : 87.8496 secs

Thi  
Engad  
Blank (us  
of

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 26, 2003, 18:30:26 ; Search time 313.496 Seconds

(without alignments)  
781.669 Million cell updates/sec

Title: US-09-696-872-23

Perfect score: 584  
Sequence: 1 MRVITGLLALAVCSAAK.....PKPQKPEPGTGSSEKDEL 109

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xjh  
-Q=/cgn2\_1/USPTO.epool/US09696872/runat\_26082003\_151139\_3337/app\_query.fasta.1.462  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62  
-TRANS=human40.cdd -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THE MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09696872@cgn\_1\_1\_291@runat\_26082003\_151139\_3337  
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSELOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.\*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
16:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
17:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Match Length	DB ID	Description
1	262.5	44.9	755	US-09-919-603-4
				Sequence 4, Appl1

2	262.5	44.9	925	10	US-09-919-603-6	Sequence 6, Appl1
3	262.5	44.9	2439	10	US-09-954-531-140	Sequence 140, App
4	262.5	44.9	2439	10	US-09-954-531-359	Sequence 359, App
5	262.5	44.9	2439	11	US-09-918-624B-49	Sequence 49, Appl1
6	262.5	44.9	2439	12	US-10-301-822-40	Sequence 40, Appl1
7	262.5	44.9	2439	14	US-10-177-293-71	Sequence 71, Appl1
8	140	24.0	537	14	US-10-154-971-21	Sequence 21, Appl1
9	139	23.8	912	14	US-10-154-971-23	Sequence 23, Appl1
10	138.5	23.7	320	9	US-09-815-343-1091	Sequence 1091, Ap
11	126	21.6	1734	14	US-10-156-761-4994	Sequence 4994, Ap
12	126	21.6	9025608	14	US-10-156-761-1	Sequence 1, Appl1
13	122	20.9	521	9	US-09-735-705-324	Sequence 324, App
14	122	20.9	521	10	US-09-850-716A-324	Sequence 324, App
15	122	20.9	521	10	US-09-897-778-324	Sequence 324, App
16	122	20.9	521	12	US-10-117-982-324	Sequence 324, App
17	122	20.9	1082	9	US-09-925-301-441	Sequence 441, App
18	122	20.9	1412	12	US-10-301-822-154	Sequence 154, App
19	122	20.9	2704	14	US-10-106-698-1371	Sequence 1371, Ap
20	122	20.9	3747	13	US-10-071-766-33	Sequence 33, Appl1
21	121	20.7	299	9	US-09-294-093B-2226	Sequence 2226, Ap
22	118	20.2	17136	14	US-10-114-170-158	Sequence 158, App
23	116	19.9	5184	14	US-10-260-715-7	Sequence 7, Appl1
24	116	19.9	88421	10	US-09-976-059-1	Sequence 1, Appl1
25	115	19.7	569	10	US-09-881-752A-117	Sequence 117, App
26	115	19.7	958	10	US-09-881-752A-227	Sequence 227, App
27	113.5	19.4	9025608	14	US-10-156-761-1	Sequence 1, Appl1
28	113	19.3	372	13	US-10-079-623-187	Sequence 187, App
29	112.5	19.3	2276	12	US-10-137-870-9	Sequence 9, Appl1
30	112.5	19.3	2276	12	US-10-140-018-9	Sequence 9, Appl1
31	112.5	19.3	2276	12	US-10-140-021-9	Sequence 9, Appl1
32	112.5	19.3	2276	12	US-10-140-274-9	Sequence 9, Appl1
33	112.5	19.3	2276	12	US-10-140-274-9	Sequence 9, Appl1
34	112.5	19.3	2276	12	US-10-140-807-9	Sequence 9, Appl1
35	112.5	19.3	2276	12	US-10-140-922-9	Sequence 9, Appl1
36	112.5	19.3	2276	12	US-10-140-924-9	Sequence 9, Appl1
37	112.5	19.3	2276	12	US-10-140-926-9	Sequence 9, Appl1
38	112.5	19.3	2276	12	US-10-141-698-9	Sequence 9, Appl1
39	112.5	19.3	2276	12	US-10-141-702-9	Sequence 9, Appl1
40	112.5	19.3	2276	12	US-10-141-704-9	Sequence 9, Appl1
41	112.5	19.3	2276	12	US-10-142-421-9	Sequence 9, Appl1
42	112.5	19.3	2276	12	US-10-142-432-9	Sequence 9, Appl1
43	112.5	19.3	2276	12	US-10-142-767-9	Sequence 9, Appl1
44	112.5	19.3	2276	12	US-10-143-033-9	Sequence 9, Appl1
45	112.5	19.3	2276	12	US-10-144-994-9	Sequence 9, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-919-603-4  
Sequence 4, Application US/09919603  
Patent No. US20020137679A1  
GENERAL INFORMATION:  
APPLICANT: Lawyer, John W.  
TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP  
FILE REFERENCE: 1440.1033-007  
CURRENT APPLICATION NUMBER: US/09/919.603  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: PCT/US00/02482  
PRIOR FILING DATE: 2000-02-01  
PRIOR APPLICATION NUMBER: 60/118,053  
PRIOR FILING DATE: 1999-02-01  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 755  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fusion gene  
US-09-919-603-4





```

Sequence 359, Application US/09954531
Patent No. US20020165180A1
GENERAL INFORMATION:
APPLICANT: Weaver, Zoe
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
FILE OF INVENTION: Gene Sets
FILE REFERENCE: 689290-77
CURRENT APPLICATION NUMBER: US/09/954,531
CURRENT FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: US/60/233,133
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,009
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,034
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,509
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US/60/234,567
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 1392
SOFTWARE: PatentIn version 3.0
SEQ ID NO 359
LENGTH: 2439
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-531-359

Alignment Scores:
Pred. No.: 4,97e-22 Length: 2439
Score: 262.50 Matches: 58
Percent Similarity: 76.47% Conservative: 7
Best Local Similarity: 68.24% Mismatches: 16
Query Match: 44.95% Indels: 5
DB: 10 Gaps: 1

US-09-636-872-23 (1-109) x US-09-954-531-359 (1-2439)
QY 8 LeuDeuAlaLeuAlaAlaValCySerAlaAlaLysIysGlySerSerLeuGlyGlyAsp 27
Db 53 CTGCTCACCCCTGGCTGCTGCTGCGGCGCTCGGACAGGCGCCAGACCCGCTTGCGC----- 106
QY 28 CysCySerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaLeu 47
Db 107 -----TCAGACCTCCGGGCGCCGACAGATGCTTCGGGAACTGCAGGAACAACACGCGCGCTG 160
QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
Db 161 CAGGACCTGGCGGAGCTGCTGCGGACAGGTCAGGAGATCAAGTCTTGAAAAAACAACG 220
QY 68 ValMetGluCyAspAlaCyGlyProGlnProGlnProGlnProGlnProGln 87
Db 221 GTAGTAGAGTGATGACGCGCTGCGGAGATGACAGATCAATGACGAC-CGGCTTACCAGCT 279
QY 88 ProGlnProLysPro 92
Db 280 GCGGCGCCCTGCTCCA 294

RESULT 5
US-09-918-624B-49
Sequence 49, Application US/09918624B
Publication No. US20030113720A1
GENERAL INFORMATION:
APPLICANT: Schenke, Xiao Min
APPLICANT: Sornase, Thierry
TITLE OF INVENTION: CDNAS EXPRESSED IN ADIPOCYTE DIFFERENTIATION.
FILE REFERENCE: PA-0033 US
CURRENT APPLICATION NUMBER: US/09/918,624B
CURRENT FILING DATE: 2002-12-03
PRIOR APPLICATION NUMBER: 60/222,470
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PERL Program
SEQ ID NO 49

```

```

LENGTH: 2439
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Genbank ID No. US20030113720A1 9602449
US-09-918-624B-49

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:

          4,97e-22      Length:      2439
          262.50       Matches:        58
          76.47%       Conservative:   7
          68.24%       Mismatches:     16
          44.95%       Indels:         5
                   Gaps:              1

US-09-696-872-23 (1-109) x US-09-918-624B-49 (1-2439)

QY      8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysGlySerSerLeuGlyGlyAsp 27
Db      53 CTGTCTACCCCTGGTGCCTCGGGCGCTCCGAGACAGGGCCAGAGCCCGTGggc----- 106
QY      28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnIuThrAsnAlaLeu 47
Db      107 -----TCAGACTCTGCCCGCCGACAGATCTTCTGGGAATTGACGAAACCAACGCCGCTG 160
QY      48 GluAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
Db      161 CAGGACCTGGGGGAGCTGGCTGCGGCGACAGCTACGAGAGATCACGTTCTGAAAAAACAG 220
QY      68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGln 87
Db      221 GTGATGAGGTGACGCGCTCGGGATGACAGCAGTAGTACGCAC-CGGCTTACCAGCGT 279
QY      88 ProGlnProLysPro 92
Db      280 GCGGCCCCCTGCTCCA 294

RESULT 6
US-10-301-822-40
Sequence 40, Application US/10301822
Publication No. US20030148410A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Berger, Allison
APPLICANT: Guillemette, Tracy L.
APPLICANT: Kametkar, Shubhangi
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Tribodeau, Stephen N.
APPLICANT: Burgart, Lawrence J.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF COLON CANCER
FILE REFERENCE: MEMO1-029P2RNM
CURRENT APPLICATION NUMBER: US/10/301,822
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US 60/339,971
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 60/361,978
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/381,988
PRIOR FILING DATE: 2002-05-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 40
LENGTH: 2439
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (26)...(2299)
US-10-301-822-40
```

Alignment Scores:			
Pred. No.:	4,97e-22	length:	2439
Score:	262.50	Matches:	58
Percent Similarity:	76.47%	Conservative:	7
Best Local Similarity:	68.24%	Mismatches:	16
Query Match:	44.95%	Indels:	5
DB:	12	Gaps:	1
US-09-696-872-23 (1-109) x US-10-301-822-40 (1-2439)			
QY	8 LeuLeuAlaLeuAlaAlaValCySerAlaAlaLyLysGlySerSerLeuGlyGlyAap	27	
Db	53 CTGCTCACCTGCTGCTGCTCGGCGGCGTCCGACAGGAGCCAGCCGTTGGCC-----	106	
QY	28 CyCysSerSerLeuGlyProGlnMetLeuArgGlnLeuGlnGlnThrMetAlaAlaLeu	47	
Db	107 -----TCACACCTGGGCGCCGACAGATGCTTCGGGAACTGACAGAAACCAACGGGGCTG	160	
QY	48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGlnIleThrPheLeuYsaenthr	67	
Db	161 CAGACCTGGGAGACTGCTGCGCGACAGCTACGGAGATCACGTTCTGAAAAACAG	220	
QY	68 ValMetGlnCyAspAlaCyGlyProGlnProGlnProLeuProLeuProGlnProGln	87	
Db	221 GTGATGAGAGTGTAACGCGTGGGATGCAGACTCAGTACGACAC-CGGCCTACCCAGCGT	279	
QY	88 ProGlnProLeuPro 92		
Db	280 GCGGCCCCCTGCTCCA 294		
RESULT 7			
US-10-177-293-71			
Sequence 71, Application US/10177293			
Publication No US20030124128A1			
GENERAL INFORMATION:			
APPLICANT: Lillie, James			
APPLICANT: Glatz, Karen			
APPLICANT: Zhao, Xumei			
APPLICANT: Gannavarpu, Manjula			
APPLICANT: Kametkar, Shubhang			
APPLICANT: Mertens, Maureen			
APPLICANT: Myer, Vic			
APPLICANT: Wang, Youzhen			
APPLICANT: Xu, Yongyao			
APPLICANT: Hoersch, Sebastian			
APPLICANT: Monahan, John			
APPLICANT: Meyers, Rachel E.			
APPLICANT: Bast Jr., Robert C.			
APPLICANT: Hortobaeyi, Gabriel N.			
APPLICANT: Puztai, Lajos			
APPLICANT: Meric, Funda			
APPLICANT: Sahin, Aysegül			
APPLICANT: Mills, Gordon B.			
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT			
TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER			
FILE REFERENCE: MRI-038			
CURRENT APPLICATION NUMBER: US/10/177,293			
CURRENT FILING DATE: 2002-06-21			
PRIOR APPLICATION NUMBER: US 60/299,887			
PRIOR FILING DATE: 2001-06-21			
PRIOR APPLICATION NUMBER: US 60/301,572			
PRIOR FILING DATE: 2001-06-27			
PRIOR APPLICATION NUMBER: US 60/306,501			
PRIOR FILING DATE: 2001-07-18			
PRIOR APPLICATION NUMBER: US 60/325,002			
PRIOR FILING DATE: 2001-09-25			
PRIOR APPLICATION NUMBER: US 60/362,585			
PRIOR FILING DATE: 2002-03-05			
PRIOR APPLICATION NUMBER: US 60/xxx,xxx			
PRIOR FILING DATE: 2002-05-14			
NUMBER OF SEQ ID NOS: 506			
SOFTWARE: FastSeq for Windows Version 4.0			

```

/ SEQ ID NO 71
/ LENGTH: 2439
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-177-293-71

Alignment Scores:
Score: 4,97e-22      Length: 2439
Percent Similarity: 76.47%      Matches: 58
Best Local Similarity: 68.24%      Conservative: 7
Query Match: 44.95%      Mismatches: 16
DB: 14      Indels: 5
      Gaps: 1

US-09-696-872-23 (1-109) x US-10-177-293-71 (1-2439)

QY      8  LeuLeuAlaLeuAlaAlaValCysSerxAlaAlaIaIyIaIySgIySerSerLeuGlyIaAsp 27
Db      53  CIGCTCACCTCGCTGCGCTCGGCCGCTCCGACAGGGCCAGACCCTGGCGC----- 106

QY      28  CysCysSerAspLeuGlyProGlnMetLeuArgLuuLeuGlnGluIuThrAsnaAlaLeu 47
Db      107  -----TCAGACTCTGGCGCCGACAGATGCTTCGGGAATCGACGAACCAACGGCGCGCTG 166

QY      48  GluAspValaIaAspTTrpLeuArgGlnGlnValaIaArgIuIeThrPheLeuIyAsnThr 67
Db      161  CAGGACGTCCGGAGACTGGCTGCGGACGACGATCAGGGAGATCACAGTTCGAAAAACACG 220

QY      68  ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGlnProGln 87
Db      221  GTCATGAGATGTGACGCGCTGCGGACGATGACGACGATCAGTACGACAC-CGGCTACCCAGCT 279

QY      88  ProGlnProLysPro 92
Db      280  GCGGCCCTGCTCCA 294

RESULT 8
US-10-154-971-21
/ Sequence 21, Application US/10154971
/ Publication No. US2003008074A1
/ GENERAL INFORMATION:
/ APPLICANT: Hamers, Raymond
/ TITLE OF INVENTION: VARIABLE FRAGMENTS OF IMMUNOGLOBULINS -
/ MyIdermans, Serge
/ NUMBER OF SEQUENCES: 29
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SPENCER & FRANK
/ STREET: 1100 New York Avenue, N.W., Suite 300 East
/ CITY: Washington
/ STATE: DC
/ COUNTRY: USA
/ ZIP: 20005
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/154,971
/ FILING DATE: 28-May-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/945,244
/ FILING DATE: <Unknown>
/ APPLICATION NUMBER: PCT/EP96/01725
/ FILING DATE: 25-APR-1996
/ APPLICATION NUMBER: EP 95400932.0
/ FILING DATE: 25-APR-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Gollin, Michael A.
/ REGISTRATION NUMBER: 31,957
/ REFERENCE/DOCKET NUMBER: GUPLA 0003

```

```

TELEFAX: 202-414-4040
INFORMATION FOR SEQ ID NO.: 23
SEQUENCE CHARACTERISTICS:
    LENGTH: 912 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
    NAME/KEY: CDS
    LOCATION: 1..309
SEQUENCE DESCRIPTION: SEQ ID NO: 23
US-10-154-971-23

Alignment Scores:
Pred. No.:      1.16e-07          Length:           912
Score:         139.00            Matches:           23
Percent Similarity: 100.00%       Conservative:     23
Best Local Similarity: 100.00%     Mismatches:        0
Query Match:   23.80%             Indels:            0
DB:            Gaps:               0

US-09-696-872-23 (1-109) x US-10-154-971-23 (1-912)

OY      76 ProGlnProGlnProLyPProGlnProGlnProGlnProGlnProLyPProGlnProLyS 95
Db      412 CCACCAACAACAAACCAAACCAAGCCAACCAACCAACCAACGAAGCAGAAAAACCAACCAAAA 471
OY      96 ProGluPro 98
Db      472 CCTGAACC 480

RESULT 10
US-09-815-343-1091
Sequence 1091, Application US/09815343
Patent No. US2001005596A1
GENERAL INFORMATION:
APPLICANT: Meagher, Madeleine
APPLICANT: Xu, Jiangchun
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.504
CURRENT APPLICATION NUMBER: US/09/815.343
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1091
LENGTH: 320
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(320)
OTHER INFORMATION: n = A,T,C or G
US-09-815-343-1091

Alignment Scores:
Pred. No.:      3.95e-08          Length:           320
Score:         138.50            Matches:           28
Percent Similarity: 53.62%       Conservative:     29
Best Local Similarity: 40.58%     Mismatches:       29
Query Match:   23.72%             Indels:           3
DB:            Gaps:              1

US-09-696-872-23 (1-109) x US-09-815-343-1091 (1-320)

OY      31 AspleUGlYProGlInMetLeuArgGlUeuGlInGuIurHraenaIAalaLauGlnAspVal 50
Db      60 GACTTTAAACCGCATGTTCCTGGGATCAATTGCACACTTAAACCAACTCCCGAGAGAGTG 119
OY      51 ArgAspTrpLeuArgGlnGlnValArgGluIleThrPhelEuLySAanthrValwMeGlu 70

```

```

Db      120  AAGACCTTCTGAGACAGCAGGTTAAGAAAATCATCATTTTGGCAAAACATCATGCTGAA 179
Qy      71  CybA5pAlACy6e1yProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 87
Db      180  TGGCAGGCTTGGCGGNCCTTCAAGTTTCAgTCTCCGACCACAAGCAGCGNGTCCCGG 239
Qy      88  ProGlnProLy5pProGlnProLy5pPro 96
Db      240  GCTCCCTCTGCACCCGTCAACACACNCCCA 266

RESULT 11
US-10-156-761-4994
; Sequence 4994, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMCURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4994
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1734)
US-10-156-761-4994

Alignment Scores:
Pred. No.:      8,98e-06      Length:      1734
Score:          126.00      Matches:      19
Percent Similarity: 100.00%      Conservative: 5
Best Local Similarity: 79.17%      Mismatches: 0
Query Match:      21.58%      Indels:      0
                        Gaps:      0

US-09-696-872-23 (1-109) x US-10-156-761-4994 (1-1734)

Qy      76  ProGlnProGlnProLy5pProGlnProGlnProGlnProGlnProGlnProGlnProLy5 95
Db      106  CCGCAGCCCGCAGCCGCGCAGCGCAGCGCAGCGCAGCGCAGCGCAGCCGACCCGACG 165
Qy      96  ProGlnProGln 99
Db      166  CCGCAACCCCGAC 177

RESULT 12
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMCURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761

```

```
US-09-696-872-23 (1-109) x US-09-735-705-324 (1-521)
```

```
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/key: misc_feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
```

```
Alignment Scores:
Pred. No.:          0.185           Length:      9025608
Score:             126.00         Matches:       19
Percent Similarity: 100.00%     Conservative:   5
Best Local Similarity: 79.17%    Mismatches:    0
Query Match:        21.58%      Indels:         0
DB:                 14           Gaps:          0
```

```
US-09-696-872-23 (1-109) x US-10-156-761-1 (1-9025608)
Oy              76 ProGlnProGlnProLysProGlnProGlnProGlnProGlnProGlnProLys 95
Db              6065692 CCGCAGCCGCCGACGGCCGCACGCCGACGCCGACGCCGACGCCGACGCCGAG 6065751
                  |||::|||::|
Oy               96 ProGlnProGlu 99
Db              6065752 CCGCACCCCCGAG 6065763
                  |||::|||::|
```

```
RESULT 13
US-09-735-705-324
; Sequence 324, Application US//09735705
; Patent No. US20020052329A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kairos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hosken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.455C14
; CURRENT APPLICATION NUMBER: US//09/735,705
; NUMBER OF SEQ ID NOS: 419
; SOFTWARE: FastSeq for windows Version 3.0
; SEQ ID NO 324
; LENGTH: 521
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-735-705-324
```

```
Alignment Scores:
Pred. No.:          6.74e-06           Length:      521
Score:             122.00         Matches:       19
Percent Similarity: 95.83%     Conservative:   4
Best Local Similarity: 79.17%    Mismatches:    1
Query Match:        20.89%      Indels:         0
DB:                 9           Gaps:          0
```

```

US-09-897-778-324
Alignment Scores:
Pred. No.: 6.74e-06 Length: 521
Score: 122.00 Matches: 19
Percent Similarity: 95.83% Conservative: 4
Best Local Similarity: 79.17% Mismatches: 1
Query Match: 20.89% Indels: 0
DB: 10 Gaps: 0

US-09-696-872-23 (1-109) x US-09-897-778-324 (1-521)
QY 76 ProGluProGluInProGlyseProGluInProGluInProGluInProGluInProGly 95
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
Db 189 CCTCAGCCTCAGCCCGCAACCCCGCCCAATCACAACCCCGCCTCAGCCCGCAACCCAAAG 248
CCTCAGCCTCAGCCCGCAACCCCGCCCAATCACAACCCCGCCTCAGCCCGCAACCCAAAG 248
QY 96 ProGluProGlu 99
|||||.....:|||||.....:
Db 249 CCTCAGCCCGCAG 260
CCTCAGCCCGCAG 260

Search completed: August 26, 2003, 21:02:49
Job time : 1541.5 secs

```

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 26, 2003, 16:41:45 ; Search time 2940.11 Seconds  
(without alignments)  
901.052 Million cell updates/sec

Title: US-09-696-872-23  
Perfect score: 584  
Sequence: 1 MRYMIGLALAAVCSNAKK.....PKPQKPEPGTGSSEKDEL 109

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-Q=/cgr2\_1/USPRO.spool/US09696872/runat\_26082003\_151137\_3236/app.query.fasta.1.462  
-DB=EST -QPMT=fastap -SUFFIX=xt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=Dlosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09696872.@CGN 1.13396.@runat\_26082003\_151137\_3236 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGESQDRY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlm:\*  
4: em\_estlm:\*  
5: em\_estoy:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vtc:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	259.5	44.4	518	10 BE809033	BE809033 214373 MA
2	259.5	44.4	604	9 AV612410	AV612410 AV612410
3	249	42.6	496	14 CB712965	CB712965 AMGNNUC:T
4	249	42.6	561	12 BG900312	BG900312 HOA37-1-C
5	248	42.5	337	13 BY350582	BY350582 BY350582
6	248	42.5	359	13 BY335264	BY335264 BY335264
7	246.5	42.2	365	12 BG900764	BG900764 HOA42-1-H
8	246	42.1	350	12 BG897877	BG897877 HOA25-1-G
9	245.5	42.0	328	9 AT751993	AT751993 cml3a10.x
10	245.5	42.0	366	12 BG897326	BG897326 HOA12-1-E
11	245.5	42.0	370	12 BG897214	BG897214 HOA33-1-F
12	245.5	42.0	371	12 BG898406	BG898406 HOA8-1-F1
13	245.5	42.0	372	12 BG899727	BG899727 HOA32-1-A
14	245.5	42.0	373	12 BG896387	BG896387 HOA30-1-E
15	245.5	42.0	374	9 AM068338	AM068338 ch20g10.Y
16	245.5	42.0	375	12 BG900556	BG900556 HOA42-1-C
17	245.5	42.0	379	12 BG898027	BG898027 HOA20-1-H
18	245.5	42.0	380	12 BG896332	BG896332 HOA29-1-D
19	245.5	42.0	381	12 BG900343	BG900343 HOA37-1-F
20	245.5	42.0	383	12 BG898382	BG898382 HOA8-1-F4
21	245.5	42.0	384	12 BG896503	BG896503 HOA30-1-C
22	245.5	42.0	384	12 BG898936	BG898936 HOA40-1-E
23	245.5	42.0	384	12 BG900459	BG900459 HOA41-1-C
24	245.5	42.0	385	12 BG897369	BG897369 HOA12-1-B
25	245.5	42.0	385	12 BG897511	BG897511 HOA11-1-A
26	245.5	42.0	385	12 BG898968	BG898968 HOA21-1-C
27	245.5	42.0	385	12 BG900465	BG900465 HOA41-1-C
28	245.5	42.0	386	12 BG898309	BG898309 HOA39-1-A
29	245.5	42.0	386	12 BG898604	BG898604 HOA19-1-G
30	245.5	42.0	386	12 BG899695	BG899695 HOA39-1-D
31	245.5	42.0	388	12 BG896715	BG896715 HOA45-1-D
32	245.5	42.0	390	12 BG897500	BG897500 HOA14-1-G
33	245.5	42.0	390	12 BG898326	BG898326 HOA55-1-C
34	245.5	42.0	390	12 BG898329	BG898329 HOA55-1-D
35	245.5	42.0	391	9 AT751616	AT751616 cml1a01.Y
36	245.5	42.0	392	12 BG898451	BG898451 HOA9-1-D1
37	245.5	42.0	393	9 AM067829	AM067829 cml9f09.x
38	245.5	42.0	393	12 BG898353	BG898353 HOA55-1-D
39	245.5	42.0	394	12 BG900134	BG900134 HOA51-1-D
40	245.5	42.0	395	12 BG899861	BG899861 HOA40-1-C
41	245.5	42.0	397	12 BG898320	BG898320 HOA55-1-C
42	245.5	42.0	399	12 BG897459	BG897459 HOA13-1-H
43	245.5	42.0	399	12 BG899824	BG899824 HOA40-1-F
44	245.5	42.0	401	12 BG896555	BG896555 HOA47-1-D
45	245.5	42.0	401	12 BG901017	BG901017 HOA52-1-G

## ALIGNMENTS

RESULT 1  
LOCUS BE809033  
DEFINITION 214373 MAREC 280V Bos taurus cdna 5', mRNA sequence.  
ACCESSION BE809033  
VERSION BE809033.1 GI:10240145  
KEYWORDS EST.  
SOURCE Bos taurus (cow)  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; 1 (bases 1 to 518)





```

Db      298 CCGCTAAGCAGTGTGCGCGGCTTCTTCCCGAGTGGCTTGACCGAGACAGCC 357
Qy      95 -----lysProgluProgluLy---ThrglySer 103
Db      358 AACGCGCGCGCTCGGAGCCTGCGCCGAGAGTTTCAACAGCAAC 402

RESULT 3
LOCUS   CB712965
DEFINITION  CB712965 496 bp mRNA linear EST 10-APR-2003
            AMGNNUC:TRGS2-00012-H5-A trgs2 (10306) Rattus norvegicus cDNA clone
            trgs2-00012-H5 5', mRNA sequence.
ACCESSION  CB712965
VERSION    CB712965
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 496)
AUTHORS   Amgen EST Program.
TITLE     Amgen Rat EST Program
JOURNAL   Unpublished
COMMENT    Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00012 row: h column: 5.
FEATURES
    source
        1..496
            /organism="Rattus norvegicus"
            /mol_type="mRNA"
            /db_xref="taxon:10116"
            /clone="trgs2-00012-H5"
            /tissue_type="muscle"
            /clone_lib="trgs2 (10306)"
            /note="Vector: C6KFP7L; Site_1: SalI; Site_2: NotI; Rat
            gastrinemiaus and soleus muscle"
BASE COUNT  76 a 147 c 129 g 75 t 69 others
ORIGIN
    Alignment Scores:
    Pred. No.: 6.94e-14 Length: 496
    Score: 249.00 Matches: 55
    Percent Similarity: 71.26% Conservative: 7
    Best Local Similarity: 63.22% Mismatches: 21
    Query Match: 42.64% Indels: 4
    DB: 14 Gaps: 1
US-09-696-872-23 (1-109) x CB712965 (1-496)
Qy      8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysIysGlySerSerLeuGlyGlyAsp 27
    ::::::::::::::::::::
Db      94 GTGCTCGCGCTGCGCTGCTGCGGAGCTACCGGAGGCGCCAGATCCCGCTGGGTGGA--- 150
Qy      28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
    ::::::::::::::::::::
Db      151 -----GACCTAGCCCCACAGATGCTTGAGAACTCCGAGAGACTTAATGCGCGCTG 201
Qy      48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
    ::::::::::::::::::::
Db      202 CAAGACGTGAGAGAGCTTTCGACAGCAGGTCAAGAGATCACTTCTTAAGATACG 261
Qy      68 ValMetGluCysAspAlaCysGlyProGlnProGlnProLysProGlnProGlnProGln 87
    ::::::::::::::::::::
Db      262 GTGATGAGATGTGACGCTTGCGGAGATGACGCCGACGACCCCGCGTCTGAGGTGCGG 321
Qy      88 ProGlnProLysProGlnPro 94
    ::::::::::::::::::::
Db      322 CCAGTGCGCTCTGCGGACCC 342

RESULT 4

```

```

BG900312
LOCUS   BG900312 561 bp mRNA linear EST 06-NOV-2001
DEFINITION  BG900312-1-C8 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
            mRNA sequence.
ACCESSION  BG900312
VERSION    BG900312.1 GI:14310561
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 561)
AUTHORS   Kumar,S., Connor,J.R., Dadds,R.A., Halsey,W., Van Horn,M., Mao,J.,
            Sathie,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
            Lark,M.W.
TITLE     Identification and initial characterization of 5000 expressed
            sequenced tags (ESTs) each from adult human normal and
            osteoarthritic cartilage cDNA libraries
JOURNAL   Osteoarthr. Cartil. 9 (7), 641-653 (2001)
MEDLINE   21482651
PUBMED    11597177
COMMENT    Contact: Sanjay Kumar
            UW2109
            GlaxoSmithKline
            709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
            Tel: 610-270-7245
            Fax: 610-270-5598
            Email: sanjay_kumar-10@sk.com
            Seq primer: T7.
FEATURES
    source
        1..561
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /tissue_type="cartilage"
            /lab_host="E.coli DH10 B"
            /clone_lib="HOA (Human Osteoarthritic Cartilage)"
            /note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
            Directional"
BASE COUNT  92 a 212 c 171 g 85 t 1 others
ORIGIN
    Alignment Scores:
    Pred. No.: 8e-14 Length: 561
    Score: 249.00 Matches: 57
    Percent Similarity: 73.03% Conservative: 8
    Best Local Similarity: 64.04% Mismatches: 20
    Query Match: 42.64% Indels: 5
    DB: 12 Gaps: 1
US-09-696-872-23 (1-109) x BG900312 (1-561)
Qy      8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLysIysGlySerSerLeuGlyGlyAsp 27
    ::::::::::::::::::::
Db      137 CTGCTCACCTGAGCTGCTGCGGAGCTGCGGAGCAGGAGCCAGAGCCGTTGGGC----- 190
Qy      28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGluThrAsnAlaAlaLeu 47
    ::::::::::::::::::::
Db      191 -----TCAGACCTGGGCGCCGAGATGCTTCGGGAACTGAGAGAAACCAACGCGGCGCTG 244
Qy      48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 67
    ::::::::::::::::::::
Db      245 CAGGACGTGCGGAGAGCTGCTGCGGACAGAGTCAAGAGATCACTTCTTAAGATACG 304
Qy      68 ValMetGluCysAspAlaCysGlyProGlnProGlnProLysProGlnProGlnProGln 87
    ::::::::::::::::::::
Db      305 GTGATGAGTGTGACGCTGCGGAGATGACAGATCACTGATGACAC-CGCGCTACCCAGCGT 363
Qy      88 ProGlnProLysProGlnProLysPro 96
    ::::::::::::::::::::
Db      364 CGGCGCCCTGCTCCACATGCGCGCCCG 390

RESULT 5

```

BY350582 337 bp mRNA linear EST 12-DEC-2002  
 LOCUS BY350582 RIKEN full-length enriched, whole joints Mus musculus CDNA  
 DEFINITION clone L830018K02 5', mRNA sequence.  
 ACCESSION BY350582  
 VERSION BY350582.1 GI:26580070  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 337)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A.,  
 Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,  
 Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chochia, C., Corbani,  
 L.E., Cousine, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,  
 A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,  
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,  
 Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M.,  
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,  
 P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,  
 H., Nagashima, T., Numata, K., Okido, T., Pavan, M.J., Pertea, G.,  
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ring,  
 Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,  
 M., Shimada, K., Sultana, R., Takenaki, Y., Taylor, M.S., Teasdale,  
 R.D., Tomita, M., Verardo, R., Wagner, L., Wahlested, C., Wang, Y.,  
 Watanabe, Y., Wells, C., Wilming, L.G., Wymshaw-Boris, A., Yanagisawa,  
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
 Carinini, P., Hayatsu, N., Hirozane-Kikukawa, T., Kono, H., Nakamura,  
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Atzawa, K.,  
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,  
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sakai, D., Shibata,  
 K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterson, R., Lander,  
 E.S., Rogers, J., Birney, E. and Hayashizaki, Y.  
 Analyses of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
 22354683  
 12466851  
 CONTACT Yoshitake Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsr.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Atzawa, K., Akimura, T., Arakawa, T., Carinini, P., Fukuda, S., Hirozane,  
 T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A.,  
 Murata, M., Nakamura, M., Nomura, K., Nunazaki, R., Ohno, M., Sakai, K.,  
 Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
 M., Waki, K., Watanabe, A., Yamashita, M. and Hayashizaki, Y. Direct  
 Submission  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN  
 Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.  
 Tissues were provided by Vassilis Aidinis ( Biomedical Sciences  
 Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.  
 Fleming street 16772 Vari, Greece ) whose assistance we gratefully  
 acknowledge  
 Please visit our web site (http://genome.gsc.riken.go.jp) for  
 further details.  
 FEATURES  
 source  
 1. .337  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="L830018K02"  
 /tissue\_type="whole joints"  
 /clone\_id="RIKEN full-length enriched, whole joints"  
 BASE COUNT 63 a 108 c 108 g 58 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 5.5e-14 Length: 337  
 Score: 248.00 Matches: 58  
 Percent Similarity: 65.00% Conservative: 7  
 Best Local Similarity: 58.00% Mismatches: 31  
 Query Match: 42.47% Indels: 4  
 Ds: 13 Gaps: 1  
 US-09-696-872-23 (1-109) x BY350582 (1-337)  
 QY 8 LeuLeuAlaLeuAlaAlaValCysSerAlaAlaLeuValGlySerSerLeuGlyGlyAsp 27  
 Db 47 GAGCGCGCGCGTATCTGCGGCGCAGACGCCAGATCCGCTGGGAGG-- 103  
 QY 28 CysCysSerAspLeuGlyProGlnMetLeuArgGluLeuGlnGlnGlnGlnGlnGlnGln 47  
 Db 104 -----GACCTGGCGCCACACATGCTGCGAGAACTTACAGAGACTAATGCGCGCTG 154  
 QY 48 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluLeuTrpPheLeuValAsnThr 67  
 Db 155 CAAAGCGGAGAGAGCGCTTGGCGACAGAGTCAAGAGATCACTTCTGAGAAATACG 214  
 QY 68 ValMetGluCysAspAlaCysGlyProGlnProGlnProGlnProGlnProGlnProGln 87  
 Db 215 GTGATGGAATGTGATGCTGCGGATGAGCGCCACGACCTCAGGCGCTGAGCGTGG 274  
 QY 88 ProGlnProLysProGlnProLysProGlnProGlnGlyThrGlySerSerGluLysAsp 107  
 Db 275 CCAAGCGCGCGCTGCGACCGCGCTGCTCCCGGCGAGTCTGCTCGAGACAGAT 334  
 RESULT 6  
 LOCUS BY353264 359 bp mRNA linear EST 11-DEC-2002  
 DEFINITION BY353264 RIKEN full-length enriched, synovial fibroblasts Mus  
 musculus CDNA clone L130042013 5', mRNA sequence.  
 ACCESSION BY353264  
 VERSION BY353264.1 GI:26529093  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 359)  
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,  
 Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.,  
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,  
 Gojobori, T., Baldarelli, R., Hill, D.P., Bull, C., Hume, D.A.,  
 Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,  
 Beisel, K.W., Blake, J.A., Brad, D., Brusic, V., Chochia, C., Corbani,  
 L.E., Cousine, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,  
 A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,  
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,  
 Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M.,  
 King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,  
 P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT

H., Nagashima, T., Numata, K., Okido, T., Pavan, M. J., Perce, G.,  
Pesole, G., Petrovsky, N., Pillai, R., Ponting, J. U., Qi, D.,  
Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,  
B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou,  
M., Shindana, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,  
R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,  
Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yangisawa,  
M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,  
Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura,  
M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,  
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,  
Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,  
K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,  
E. S., Rogers, J., Birney, E. and Hayashizaki, Y.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nucleic Acids Res. 31: 1234-1243 (2003)

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gs.c.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane,  
T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A.,  
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,  
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,  
M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct  
Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Vassilis Aidinis ( Biomedical Sciences  
Research Center, "A. Fleming" Institute of Immunology 14-16 Al.  
Fleming street 16672 Vathi, Greece ) whose assistance we gratefully  
acknowledge.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

FEATURES  
SOURCE

Location/Qualifiers  
1. .359  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="U130042033"  
/cell\_type="synovial fibroblasts"  
/clone\_id="RIKEN full-length enriched, synovial  
fibroblasts"  
BASE COUNT 63 a 119 c 117 g 60 t  
ORIGIN

Alignment Scores:  
Pred. No.: 5.92e-14 Length: 359  
Score: 248.00 Matches: 58  
Percent Similarity: 65.66% Conservative: 7  
Best Local Similarity: 58.59% Mismatches: 30

Query Match: 42.47% Indels: 4  
DB: 13 Gaps: 1  
US-09-696-872-23 (1-109) x BY355264 (1-359)  
QY 8 leuleuAlaleuAlaAlaValCySserAlaAlaLysLysGlySerSerLeuGlyAsp 27  
D 47 GTGCTGCCTCCCTGATCTCTGCGGACACAGCCAGGCCAGATCCCGCTGGTGGA--- 103  
QY 28 CyScSserSerLeuGlyProGlnMetLeuAlaGluLeuGlnGlnGlnGlnGlnGln 47  
D 104 -----CACTCGGCCCAAGATCTCGAGAACTTCAAGAGACTTAATCGCGCTG 154  
QY 48 GlnAspValArgAspTdpLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 67  
D 155 CAGACCTGAGAGAGCTCTTCCAGACAGCTCAAGAGATCACTTCTGAGAAATGCG 214  
QY 68 ValMetGluCySAspAlaCyGlyProGlnProGlnProGlnProGlnProGln 87  
D 215 GTGATGAAATGTGATGCTTCCGGAATGACAGCCGACGACTCAGGCTGAGGCTCGG 274  
QY 88 ProGlnProLysProGlnProLysProGlnProGlnProGlnProGlnProGln 106  
D 275 CCAATGCGCTGCGGACCGGCTCTGCTCCCGCGGCTGCTGCTCGGAGAA 331  
RESULT 7  
BG900764 365 bp mRNA linear EST 06-NOV-2001  
LOCUS BG900764 365 bp mRNA linear EST 06-NOV-2001  
DEFINITION HNA4-2-1-H5 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,  
mRNA sequence.  
ACCESSION BG900764  
VERSION BG900764.1 GI:14311013  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 365)  
AUTHORS Kumar, S., Connor, U.R., Dodds, R.A., Halsey, W., Van Horn, M., Mao, J.,  
Saehe, G., Mui, P., Agarwal, P., Badger, A.M., Lee, J.C., Gowen, M. and  
Lark, M.W.  
Identification and initial characterization of 5000 expressed  
sequenced tags (ESTs) each from adult human normal and  
osteoarthritic cartilage cDNA libraries  
Osteoarthr. Cartil. 9 (7), 641-653 (2001)  
JOURNAL MEDLINE 21482651  
PUBMED 11597177  
COMMENT Contact: Sanjay Kumar  
UN2109  
GlaxoSmithKline  
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA  
Tel: 610-270-7245  
Fax: 610-270-5598  
Email: sanjay.kumar-1@sk.com  
Seq primer: T7.  
FEATURES  
SOURCE  
Location/Qualifiers  
1. .365  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="cartilage"  
/lab\_host="E.coli DH10 B"  
/clone\_id="HOA (Human Osteoarthritic Cartilage)"  
/note="Vector: pSPORT 1; Site\_1: SalI; Site\_2: NotI;  
Directional"  
BASE COUNT 65 a 123 c 121 g 56 t  
ORIGIN

Alignment Scores:  
Pred. No.: 8.33e-14 Length: 365  
Score: 246.50 Matches: 55  
Percent Similarity: 75.61% Conservative: 7  
Best Local Similarity: 67.07% Mismatches: 9





DB:	12	Gap:	2
US-09-696-872-23 (1-109) x BG897214 (1-370)			
OY	8	LeuLeuAaLaLeuAaLaAaLaValCySerAaLaAaLaAaLySgLySerSerLeuGlyGlyAsp	27
DB	137	CTGCTCACCCCTGGCTGGCTGGCTGGCCGCGCCGACACAGGGCCAGAGCCCTTGGG-----	190
OY	28	CysCysSerAaPLeuGlyProGlnMetLeuAaArgGluLeuGlnGluThrAsnAaLaLeu	47
DB	191	-----TCAGACCTGGGCGCCGAGATGCTTGGGAACTGACAGAAACCAAGCGGCGCTG	244
OY	48	GlnAaPValAaAaPTrpLeuAaArgGlnGlnValAaArgGluLeuPheLeuLyAsnThr	67
DB	245	CAGACGCGCGGAGACTCTCTGGCAGAGAGTCAAGAGATCACTTCTGAAAAAACG	304
OY	68	ValMetGluCyAaPAlaCyGlyProGln-----ProGlnPro	80
DB	305	GTGATGAGACTGTGACGCTGCGGATGACAGCAGTACGACCGGCGCTACCGCCTG	364
OY	81	LysPro	82
DB	365	CGGCC	370
RESULT 12			
BS898406		371 bp	mRNA
LOCUS		HOA8-1-F10 HOA (Human Osteoarthritic Cartilage)	linear EST 06-NOV-2001
DEFINITION			mRNA sequence.
ACCESSION	BS898406		
VERSION	BS898406.1	GI:14308655	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
REFERENCE	1 (bases 1 to 371)		
AUTHORS	Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,		
	Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and		
	Lack,M.W.		
TITLE	Identification and initial characterization of 5000 expressed		
	sequenced tags (ESTs) each from adult human normal and		
JOURNAL	osteoarthritic cartilage cDNA libraries		
MEDLINE	21482651		
PUBMED	11597177		
COMMENT	Contact: Sanjay Kumar		
	UM2109		
	GlaxoSmithKline		
	709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA		
	Tel: 610-270-7245		
	Fax: 610-270-5598		
	Email: sanjay.kumar-1@glk.com		
	Seq primer: 17		
FEATURES			
source	Location/Qualifiers		
	1..371		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="caxon:9606"		
	/tissue_type="cartilage"		
	/lab_host="E.coli DH10 B"		
	/clone_lib="HOA (Human Osteoarthritic Cartilage)"		
	/note="Vector: pSPORT 1, Site_1: SalI, Site_2: NotI;		
	Directional"		
BASE COUNT	65 a 126 c 118 g 62 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,05e-13	Length:	371
Score:	245.50	Matches:	56
Percent Similarity:	75.29%	Conservative:	8
Best Local Similarity:	65.88%	Mismatches:	17
Query Match:	42.04%	Indels:	5

DB:	12	Gaps:	1
US-09-696-872-23 (1-109) x BG898406 (1-371)			
OY	8	LeutealatalaalaaValaCySerSeraAlaIalaIyalyGlySerSerleuglYgIyAap	27
Db	105	CTGTCTACACCTGGCTGCCTGCCCTCGAGCCGACAGAGGCCAGAACCCGTTGGC-----	158
OY	28	CysCySerApPleuglYProglMetleuarargluleungInglutHraAsnaAlaleu	47
Db	159	-----TCAGACTCGGGCCCCCAATGCTTGCGGAACATGCGAAACCAGACCGCGCGCTG	212
OY	48	GluepVlaIagAAPTPlleaArglIglnAlaArgluileTHrpeleuIeuaAnthr	67
Db	213	CAGACCGTGGGGAGCTGCTCGGAGAGGTCAGGGAGATCACGTTCTTAATAAACG	272
OY	68	ValmetGIuCyaAPAlaCYeGIYProglInProglInProlySProglInProglIn	87
Db	213	GTGATGAGGTGACCTGCGCTGGGATGACAGACATCAGTACGCAC-CGGCTTACCAGCGT	331
OY	88	ProglInProlysPro	92
Db	332	CGCGCCCTGCTCCA	346
RESULT 13			
BG899727		372 bp	mRNA linear EST 06-NOV-2001
LOCUS	HOA32-1-A8 HOA (Human Osteoarthritic Cartilage)		Homo sapiens cDNA,
DEFINITION	mRNA sequence.		
VERSION	BG899727		
KEYWORDS	BG899727.1 GI:14309992		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiade; Homo. 1 (bases 1 to 372) Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Sathe,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lack,M.W. Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritis cartilage CDNA libraries Osteoarthr. Cartil. 9 (7), 641-653 (2001) 21482651 11597177 Contact: Sanjay Kumar UM2109 GlaxoSmithKline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel.: 610-270-7245 Fax: 610-270-5598 Email: sanjay.kumar-1@gsk.com Seq primer: T7.		
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
COMMENT			
FEATURES			
source	Location/Qualifiers		
	1..372		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/cissue_type="cartilage"		
	/lab_host="E.coli DH10 B"		
	/clone_id="HOA (Human Osteoarthritic Cartilage)"		
	/note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI; Directional"		
BASE COUNT	68 a	130 c	117 g 57 t
ORIGIN			
Alignment Scores:			
Pred. No.:	1.06e-13	Length:	372
Score:	245.50	Matches:	55
Percent Similarity:	74.39%	Conservative:	6
Best Local Similarity:	67.07%	Mismatches:	10
Query Match:	42.04%	Indels:	11







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 15:53:48 ; Search time 116.717 Seconds  
(without alignment)  
148.232 Million cell updates/sec

Title: US-09-696-872-23

Sequence: 1 MRYWILGLLAAVCSAAK.....PKQPKPEPGTGSSEKDEL 109

Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: A\_Geneseq\_19Jun03.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	584	100.0	109	AAV44963	KDEL receptor inh1
2	464	79.5	109	AAV44966	KDEL receptor inh1
3	434	74.3	109	AAV44965	KDEL receptor inh1
4	428.5	73.4	115	AAV44959	KDEL receptor inh1
5	422	72.3	134	AAV44967	KDEL receptor inh1
6	414.5	71.0	115	AAV44958	KDEL receptor inh1
7	371.5	63.6	109	AAV44962	KDEL receptor inh1
8	345.5	59.2	109	AAV44961	KDEL receptor inh1
9	337.5	57.8	105	AAV44960	KDEL receptor inh1

10	309.5	53.0	90	21	AAV44964
11	262.5	44.9	242	21	AAAB00040
12	262.5	44.9	300	21	AAAB00041
13	262.5	44.9	757	21	AAAB00044
14	262.5	44.9	757	23	ABU05594
15	262.5	44.9	757	24	ABR47420
16	262.5	44.9	757	24	ABU56605
17	262.5	44.9	758	23	ABH82286
18	256.5	43.9	130	22	AAV93197
19	241	41.3	46	21	AAV44947
20	211	36.1	46	21	AAV44946
21	153	26.2	1357	22	AAV79715
22	146	25.0	35	15	AAV49560
23	146	25.0	35	15	AAV49529
24	146	25.0	54	15	AAV49722
25	146	25.0	54	15	AAV49536
26	145.5	24.9	776	22	AAV79078
27	145.5	24.9	961	23	AAV56249
28	145.5	24.9	961	23	AAV56249
29	144	24.7	24	21	AAV85032
30	144	24.7	24	21	AAV85036
31	142.5	24.4	622	22	AAV80062
32	138.5	23.7	889	15	AAV56248
33	138	23.6	74	18	AAV20047
34	133	22.8	272	10	AAV3560
35	126	21.6	329	23	AAV4305
36	124	21.2	22	23	AAV80744
37	123.5	21.1	1247	22	AAV27733
38	123.5	21.1	1247	22	AAV8731
39	123	21.1	814	22	ABG04441
40	121	20.7	46	21	AAV44950
41	119	20.4	163	23	AAV4306
42	119	20.4	230	23	AAV4307
43	119	20.4	667	19	AAV48760
44	118	20.2	378	12	AAV4160
45	117	20.0	46	21	AAV44951

## ALIGNMENTS

RESULT 1	
ID	AAV44963 standard; Protein: 109 AA.
AC	AAV44963;
DT	23-MAY-2000 (first entry)
DE	KDEL receptor inhibitor protein-6.
XX	KDEL receptor inhibitor; heat shock protein; immune response;
XX	oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;
KW	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
KW	infectious disease; allergy; autoimmune disease.
XX	
OS	Chimeric - Adenovirus E3.
OS	Chimeric - Homo sapiens.
OS	Chimeric - Camelus sp.
XX	
FX	Key
FT	Peptide
FT	1..20
FT	/label= Signal peptide
FT	/note= "Derived from adenovirus E3"
FT	Domain
FT	30..75
FT	/note= "Human COMP pentamerisation domain"
FT	76..99
FT	/note= "Camel IgG linker domain"
FX	WO200006729-A1.
XX	
PD	10-FEB-2000.
XX	

KDEL receptor inh1	
Human COMP/TSP-1 c	
Human COMP/TSP-2 c	
Human cartilage ol	
Breast cancer-asso	
Breast cancer-asso	
Lung cancer-asso	
Human thrombospond	
Human polyepitide	
Human cartilage ol	
Rat cartilage olig	
Human protein SEQ	
Sequence of hinge	
Camel 2-chain Ig h	
Sequence of a pept	
Camel Ig 2-heavy c	
Human protein SEQ	
Human thrombospond	
Human thrombospond	
Target peptide #6	
Human antibody 11b	
Human protein SEQ	
Xenopus thrombospo	
Human TSP4 pentame	
Plasmodium berghei	
Sugarcane proline	
Linker peptide #2	
Human full-length	
Novel human diagno	
Human thrombospond	
Sugarcane proline	
BOP1 protein. Mus	
PRP 378. Triticum	
Xenopus thrombospo	

PF	28-JUL-1999;	99WO-US17147.	
XX			
PR	29-JUL-1998;	98US-0124671.	
FA	(SLOC ) SLOCAN KETTERING INST CANCER RES.		
XX			
XX	Rochman JE, Mayhew M, Hoe MH;		
XX			
DR	WPI; 2000-195296/17.		
DR	N-PSDB; AAZ50497.		
XX			
PT	Inhibitors of the KDEL receptor which comprises an oligomerization		
PT	domain useful for promoting secretion of proteins which are normally		
PT	retained within the cell -		
XX			
PS	Disclosure; Fig 6; 87pp; English.		
XX			
CC	The patent discloses the use of KDEL receptor inhibitor to promote		
CC	secretion of proteins that are normally retained within the cell such as		
CC	heat shock proteins by inhibiting KDEL receptor-mediated return of		
CC	protein complexes to endoplasmic reticulum. This makes the secreted heat		
CC	shock proteins more accessible to the immune system and improves immune		
CC	response to a target antigen. The inhibitor protein comprises several		
CC	subunits where each subunit comprises an oligomerisation domain and has		
CC	at its carboxy terminus a region which binds to a KDEL receptor. The		
CC	target antigen may be associated with diseases including neoplasia such		
CC	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and		
CC	astrocytoma, with defective tumour suppressor genes, oncogenes,		
CC	infectious diseases, allergy or autoimmune diseases. The present		
CC	sequence is KDEL receptor inhibitor comprising regions encoding a		
CC	cleavable signal peptide; the oligomerisation domain of human cartilage		
CC	oligomeric matrix protein (COMP) pentamerisation domain; a camel IgG		
CC	linker domain and the carboxy-terminal sequence KDEL. The subsequence		
CC	GDCG is an alteration of rat COMP which provides increased stability via		
CC	disulphide bonds.		
XX			
SO	Sequence 109 AA;		
	Query Match 100.0%; Score 584; DB 21; Length 109;		
	Best Local Similarity 100.0%; Pred. No. 7.1e-52;		
	Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 MRWMTIGLALAAVCSAAKKGSSIGDCCSDLGPMQLRELOETNALADVDVDMLRQVRE 60		
DB	1 MRWMTIGLALAAVCSAAKKGSSIGDCCSDLGPMQLRELOETNALADVDVDMLRQVRE 60		
QY	61 ITFLKNTIMECDACGPPQPPKPPQPPQPPKPPKPPPEEGTSSSEKDEL 109		
DB	61 ITFLKNTIMECDACGPPQPPKPPQPPQPPKPPKPPPEEGTSSSEKDEL 109		
RESULT 2			
AC	AAAY4966		
XX	AAAY4966 standard; Protein; 109 AA.		
XX	AAAY4966;		
DT	23-MAY-2000 (first entry)		
DE			
XX	KDEL receptor inhibitor protein-9.		
XX			
KW	KDEL receptor inhibitor; heat shock protein; immune response;		
KW	oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;		
KW	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;		
XX	infectious disease; allergy; autoimmune disease.		
XX			
OS	Chimeric - Adenovirus B3.		
OS	Chimeric - Homo sapiens.		
OS	Chimeric - Camelus sp.		
XX			
EH	Key Location/Qualifiers		
FT	1..20		
FT	/label= Signal peptide		

FT	Domain	/note= "Derived from adenovirus E3"
FT	30..75	/notes= "Human TSP4 trimerisation domain"
FT	Domain	/note= "Camel IgG linker domain"
XX	WO200006729-A1.	
XX	PD	10-FEB-2000.
XX	PZ	28-JUL-1999; 99WO-US17147.
XX	PR	29-JUL-1998; 98US-0124671.
XX	PA	(SLOK ) SLOAN KETTERING INST CANCER RES.
PI	Rochman JE, Mayhew M, Hoe MH;	
XX	WPI; 2000-195296/17.	
DR	N-PSDB; AAZ50500.	
PT	Inhibitors of the KDEL receptor which comprises an oligomerization	
PT	domain useful for promoting secretion of proteins which are normally	
XX	retained within the cell -	
XX	Disclosure; Fig 9; 87pp; English.	
CC	The patent discloses the use of KDEL receptor inhibitor to promote	
CC	secretion of proteins that are normally retained within the cell such as	
CC	heat shock proteins by inhibiting KDEL receptor-mediated return of	
CC	protein complexes to endoplasmic reticulum. This makes the secreted heat	
CC	shock proteins more accessible to the immune system and improves immune	
CC	response to a target antigen. The inhibitor protein comprises several	
CC	subunits where each subunit comprises an oligomerisation domain and has	
CC	at its carboxy terminus a region which binds to a KDEL receptor. The	
CC	target antigen may be associated with diseases including neoplasia such	
CC	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and	
CC	aeratomytoma, with defective tumour suppressor genes, oncogenes,	
CC	infectious diseases, allergy or autoimmune diseases. The present	
CC	sequence is KDEL receptor inhibitor comprising regions encoding a	
CC	cleavable signal peptide; the oligomerisation domain of human	
CC	thrombospondin 4 (TSP4) trimerisation domain; a camel IgG linker domain	
CC	and the carboxy-terminal sequence KDEL. The subsequence GDCC is an	
CC	alteration of rat cartilage oligomeric matrix protein which provides	
CC	increased stability via disulphide bonds.	
XX	Sequence 109 AA;	
SQ	Query Match 79.5%; Score 464; DB 21; Length 109;	
	Best Local Similarity 78.0%; Pred. NO. 1.2e-35;	
	Matches 85; Conservative 9; Mismatches 15; Indels 0; Gaps 0;	
OY	1 MRVITGLLALAAVCASAKKGSSLGDCSDLGFQMLRELQETNAALQDYVDWLRLQVRE 60	
Dd	1 MRYVITGLLALAAVCASAKKGSSLGDCSDLPNNRFLQGWTQLNLGLGEVADLLRRQVKE 60	
OY	61 ITPLAKTVMECDACGPQPQPKPQPPQPKPQPKPEPGTSSESKDEL 109	
Dd	61 TSFLRTIATCAQAGPPQPKPQPPQPKPQPKPEPGTSSESKDEL 109	
RESULT 3		
ID	AAY44965 standard; Protein; 109 AA.	
XX	AAY44965;	
XX	AC	23-MAY-2000 (first entry)
XX	KDEL receptor inhibitor protein-8.	
XX	KDEL receptor inhibitor; heat shock protein; immune response;	
TW	oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;	

KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
 XX infectious disease; allergy; autoimmune disease.  
 OS Chimeric - Adenovirus E3.  
 OS Chimeric - Homo sapiens.  
 OS Chimeric - Camelus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..20  
 FT /label= Signal peptide  
 FT /note= "Derived from adenovirus E3"  
 FT Domain 30..75  
 FT /note= "Human TSP 3 trimerisation domain"  
 FT Domain 76..99  
 FT /note= "Camel IgG linker domain"  
 XX  
 PN WO200006729-A1.  
 XX  
 XX 10-FEB-2000.  
 XX  
 XX 28-JUL-1999; 99WO-US17147.  
 XX PF  
 XX 29-JUL-1998; 98US-0124671.  
 XX  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 XX  
 XX Rothman JE, Mayhew M, Hoe MH;  
 XX WPI; 2000-195296/17.  
 DR N-PSDB; AAZ50499.  
 XX  
 XX Inhibitors of the KDEL receptor which comprises an oligomerization  
 PT domain useful for promoting secretion of proteins which are normally  
 PT retained within the cell -  
 XX  
 XX Disclosure; Fig 8; 87pp; English.  
 XX  
 CC The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerisation domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence is KDEL receptor inhibitor comprising regions encoding a  
 CC cleavable signal peptide; the oligomerisation domain of human  
 CC thymospondin 3 (TSP3) trimerisation domain; a camel IgG linker domain  
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an  
 CC alteration of rat cartilage oligomeric matrix protein which provides  
 CC increased stability via disulphide bonds.  
 XX  
 XX Sequence 109 AA;  
 SQ  
 Query Match 74.3%; Score 434; DB 21; Length 109;  
 Best Local Similarity 72.5%; Pred. No. 1.4e-36;  
 Matches 79; Conservative 13; Mismatches 17; Indels 0; Gaps 0;  
 QY 1 MRWIIIGLLAAVCSAKKSSIGDCCSDLGPMRLQETNALQDVADMLKQVRE 60  
 DB 1 MRWIIIGLLAAVCSAKKSSIGDCCSDLGPMRLQETNALQDVADMLKQVRE 60  
 QY 61 ITFLKNTVMCDACGPOPKPOPKPOPKPOPKPEEGTSSSEKDEL 109  
 DB 61 MSJLRNTIMECQVCGPOPKPOPKPOPKPOPKPEEGTSSSEKDEL 109  
 RESULT 4  
 AAY44959

ID AAY44959 standard; Protein; 115 AA.  
 XX  
 XX AAY44959;  
 AC  
 XX 23-MAY-2000 (first entry)  
 DT  
 XX KDEL receptor inhibitor protein-2.  
 DE  
 XX KDEL receptor inhibitor; heat shock protein; immune response;  
 KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;  
 KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
 KW infectious disease; allergy; autoimmune disease.  
 XX  
 OS Chimeric - Mus sp.  
 OS Chimeric - Camelus sp.  
 OS Chimeric - Rattus sp.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..20  
 FT /label= Signal peptide  
 FT /note= "Derived from mouse Bip"  
 FT Domain 26..30  
 FT /note= "Altered subsequence from rat COMP"  
 FT Domain 30..71  
 FT /note= "Rat COMP pentamerisation domain"  
 FT Domain 82..105  
 FT /note= "Camel IgG linker domain"  
 XX  
 XX WO200006729-A1.  
 XX  
 XX 10-FEB-2000.  
 XX  
 XX 28-JUL-1999; 99WO-US17147.  
 XX PF  
 XX 29-JUL-1998; 98US-0124671.  
 XX  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 XX  
 XX Rothman JE, Mayhew M, Hoe MH;  
 XX WPI; 2000-195296/17.  
 DR N-PSDB; AAZ50493.  
 XX  
 XX Inhibitors of the KDEL receptor which comprises an oligomerization  
 PT domain useful for promoting secretion of proteins which are normally  
 PT retained within the cell -  
 XX  
 XX Disclosure; Fig 2; 87pp; English.  
 XX  
 CC The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerisation domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence is KDEL receptor inhibitor protein comprising regions including  
 CC a cleavable signal peptide; the oligomerisation domain from rat  
 CC cartilage oligomeric matrix protein (COMP); a camel IgG linker domain  
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an  
 CC alteration of rat COMP which provides increased stability via disulphide  
 CC bonds.  
 XX  
 XX Sequence 115 AA;  
 SQ  
 Query Match 73.4%; Score 428.5; DB 21; Length 115;  
 Best Local Similarity 77.7%; Pred. No. 5.3e-36;  
 Matches 87; Conservative 3; Mismatches 5; Indels 17; Gaps 3;

```

QY      8 LLAALAVCSAAKGGSLGDCSDLGPMRLRELOETNALQDVPRDLRQVREITFLKNT 67
      11 LLLLGAV---RAEGSLGSDCC---PQMLRELOETNALQDVRLRLAQVKEITFLKNT 63
Db      68 VMECDACG-----PQPPKQPQPQPQPQPQPPEPBGTSSEKDEL 109
      64 VMECDACGMQPARTPGTSPPQPPKQPQPQPQPQPPEPBGTSSEKDEL 115

RESULT 5
AA44967
ID      AA44967 standard; Protein, 134 AA.
AC      AA44967;
DT      23-MAY-2000 (first entry)
DE      KDEL receptor inhibitor protein-10.
XX
XX      KDEL receptor inhibitor; heat shock protein; immune response;
KW      oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
KW      melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
KW      infectious disease; allergy; autoimmune disease.
OS      Chimeric - Mus sp.
OS      Chimeric - Rattus sp.
OS      Chimeric - Camelus sp.
XX
XX      Key      Location/Qualifiers
FT      Peptide      1..20
FT      /label= Signal_peptide
FT      /note= "Derived from mouse Bip"
FT      Region      23..32
FT      /label= Myc_tag_sequence
FT      Modified-site 36..38
FT      /note= "Asn is N-glycosylated"
FT      Domain      49..94
FT      /note= "Rat COMP pentamerisation domain"
FT      Domain      101..124
FT      /note= "Camel IgG linker domain"
XX
XX      WO200006729-A1.
XX
XX      10-FEB-2000.
XX
XX      28-JUL-1999; 99WO-US17147.
XX
XX      29-JUL-1998; 98US-0124671.
XX
XX      (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
XX      Roelman JE, Mayhew M, Hoe MH;
XX
XX      WPI; 2000-195296/17.
XX      N-PSDB; AAZ50501.
XX
XX      Inhibitors of the KDEL receptor which comprises an oligomerization
PT      domain useful for promoting secretion of proteins which are normally
PT      retained within the cell -
XX
XX      Disclosure; Fig 10; 87pp; English.
XX
XX      The patent discloses the use of KDEL receptor inhibitor to promote
CC      secretion of proteins that are normally retained within the cell such as
CC      heat shock proteins by inhibiting KDEL receptor-mediated return of
CC      protein complexes to endoplasmic reticulum. This makes the secreted heat
CC      shock proteins more accessible to the immune system and improves immune
CC      response to a target antigen. The inhibitor protein comprises several
CC      subunits where each subunit comprises an oligomerisation domain and has
CC      at its carboxy terminus a region which binds to a KDEL receptor. The
CC      target antigen may be associated with diseases including neoplasia such

```

```

CC      as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
CC      astrocytoma, with defective tumour suppressor genes, oncogenes,
CC      infectious diseases, allergy or autoimmune diseases. The present
CC      sequence encodes KDEL receptor inhibitor comprising regions encoding a
CC      cleavable signal peptide; a myc-tag; an N-glycosylation sequence; the
CC      oligomerisation domain of rat cartilage oligomeric matrix protein
CC      (COMP); a camel IgG linker domain and the carboxy-terminal sequence
CC      KDEL. The subsequence GDCC is an alteration of rat COMP which provides
CC      increased stability via disulphide bonds.
XX
XX      Sequence      134 AA;
SQ
XX
XX      Query Match      72.3%; Score 422; DB 21; Length 134;
XX      Best local similarity 82.8%; Pred. No. 2.9e-35;
XX      Matches 82; Conservative 2; Mismatches 1; Indels 14; Gaps 2;
QY      21 GSSLGSDCCSDLGPMRLRELOETNALQDVPRDLRQVREITFLKNTVMECDACG----- 75
      40 GSSLGSDCC---PQMLRELOETNALQDVRLRLAQVKEITFLKNTVMECDACGMQPAR 95
Db      76 -----PQPPKQPQPQPQPQPQPPEPBGTSSEKDEL 109
      96 TPGRSPQPPKQPQPQPQPQPQPPEPBGTSSEKDEL 134

RESULT 6
AA44958
ID      AA44958 standard; Protein, 115 AA.
AC      AA44958;
DT      23-MAY-2000 (first entry)
DE      KDEL receptor inhibitor protein-1.
XX
XX      KDEL receptor inhibitor; heat shock protein; immune response;
KW      oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
KW      melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
KW      infectious disease; allergy; autoimmune disease.
OS      Chimeric - Mus sp.
OS      Chimeric - Camelus sp.
OS      Chimeric - Rattus sp.
XX
XX      Key      Location/Qualifiers
FT      Peptide      1..20
FT      /label= Signal_peptide
FT      /note= "Derived from mouse Bip"
FT      Domain      26..30
FT      /note= "Subsequence from rat COMP"
FT      Domain      30..71
FT      /note= "Rat COMP pentamerisation domain"
FT      Domain      82..105
FT      /note= "Camel IgG linker domain"
XX
XX      WO200006729-A1.
XX
XX      10-FEB-2000.
XX
XX      28-JUL-1999; 99WO-US17147.
XX
XX      29-JUL-1998; 98US-0124671.
XX
XX      (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
XX      Roelman JE, Mayhew M, Hoe MH;
XX
XX      WPI; 2000-195296/17.
XX      N-PSDB; AAZ50492.
XX
XX      Inhibitors of the KDEL receptor which comprises an oligomerization
PT      domain useful for promoting secretion of proteins which are normally
PT      retained within the cell -

```

XX PS Disclosure; Fig 1; 87pp; English.

CC The patent discloses the use of KDEL receptor inhibitor to promote

CC secretion of proteins that are normally retained within the cell such as

CC heat shock proteins by inhibiting KDEL receptor-mediated return of

CC protein complexes to endoplasmic reticulum. This makes the secreted heat

CC shock proteins more accessible to the immune system and improves immune

CC response to a target antigen. The inhibitor protein comprises several

CC subunits where each subunit comprises an oligomerisation domain and has

CC at its carboxy terminus a region which binds to a KDEL receptor. The

CC target antigen may be associated with diseases including neoplasia such

CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and

CC astrocytoma, with defective tumour suppressor genes, oncogenes,

CC infectious diseases, allergy or autoimmune diseases. The present

CC sequence is KDEL receptor inhibitor protein comprising regions including

CC a cleavable signal peptide; the oligomerisation domain from rat

CC cartilage oligomeric matrix protein; a camel IgG linker domain and the

CC carboxy-terminal sequence KDEL.

XX SQ Sequence 115 AA;

Query Match 71.0%; Score 414.5; DB 21; Length 115;

Best Local Similarity 76.8%; Pred. No. 1.4e-34;

Matches 86; Conservative 3; Mismatches 6; Indels 17; Gaps 3;

Oy 8 LIALAAVCSAANKSSSLGDCSPDLGPGMLRELOETNAALQDVWDMLRQVREITFLKNT 67

Db 11 LLLLGAV---RAEGSSLOG---DLAPQMLRELOETNAALQDVWELRLQVKEITFLKNT 63

Oy 68 VMECDACG-----POPQPKPOPOPOPKPOPKPEPEGSSSEKDEL 109

Db 64 VMECDACGMPARTRPGTSPPQPKPOPKPOPKPEPEGSSSEKDEL 115

RESULT 7

AA44962

ID AAY44962 standard; Protein; 109 AA.

XX AAY44962;

AC 23-MAY-2000 (first entry)

XX DE KDEL receptor inhibitor protein-5.

XX KW KDEL receptor inhibitor; heat shock protein; immune response;

KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;

KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;

XX infectious disease; allergy; autoimmune disease.

OS Chimeric - Mus sp.

OS Chimeric - Xenopus sp.

OS Chimeric - Camelus sp.

OS Chimeric - Rattus sp.

XX Key Location/Qualifiers

FT Peptide 1..20

FT /label= Signal peptide

FT /note= "Derived from mouse B1p"

FT Domain 26..30

FT /note= "Altered subsequence from rat COMP"

FT Domain 30..75

FT /note= "Xenopus thrombospondin 4 trimerisation domain"

FT Domain 76..99

FT /note= "Camel IgG linker domain"

XX MO200006729-A1.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17147.

XX 29-JUL-1998; 98US-0124671.

XX PA (SLOC) SLOAN KETTERING INST CANCER RES.

XX PI Rothman JE, Mayhew M, Hoe MH;

XX WPI; 2000-195296/17.

DR N-PSDB; AA250496.

XX PT Inhibitors of the KDEL receptor which comprises an oligomerization

PT domain useful for promoting secretion of proteins which are normally

PT retained within the cell

XX Disclosure; Fig 5; 87pp; English.

CC The patent discloses the use of KDEL receptor inhibitor to promote

CC secretion of proteins that are normally retained within the cell such as

CC heat shock proteins by inhibiting KDEL receptor-mediated return of

CC protein complexes to endoplasmic reticulum. This makes the secreted heat

CC shock proteins more accessible to the immune system and improves immune

CC response to a target antigen. The inhibitor protein comprises several

CC subunits where each subunit comprises an oligomerisation domain and has

CC at its carboxy terminus a region which binds to a KDEL receptor. The

CC target antigen may be associated with diseases including neoplasia such

CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and

CC astrocytoma, with defective tumour suppressor genes, oncogenes,

CC infectious diseases, allergy or autoimmune diseases. The present

CC sequence is KDEL receptor inhibitor protein comprising regions encoding a

CC cleavable signal peptide; the oligomerisation domain of Xenopus

CC thrombospondin 4 (TSP4) trimerisation domain including an additional

CC subsequence; a camel IgG linker domain and the carboxy-terminal sequence

CC KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric

CC matrix protein which provides increased stability via disulphide bonds.

XX SQ Sequence 109 AA;

Query Match 63.6%; Score 371.5; DB 21; Length 109;

Best Local Similarity 66.7%; Pred. No. 3.2e-30;

Matches 68; Conservative 11; Mismatches 18; Indels 3; Gaps 1;

Oy 8 LIALAAVCSAANKSSSLGDCSPDLGPGMLRELOETNAALQDVWDMLRQVREITFLKNT 67

Db 11 LLLLGAV---RAEGSSSLGDCSPDVSRLIGQITQNMQMGELRDVWRQVKEITFLKNT 67

Oy 68 VMECDACGPOPOPKPOPOPKPOPKPEPEGSSSEKDEL 109

Db 68 IAEQACGPOPOPKPOPOPKPOPKPEPEGSSSEKDEL 109

RESULT 8

AA44961

ID AAY44961 standard; Protein; 109 AA.

XX AAY44961;

AC 23-MAY-2000 (first entry)

XX DE KDEL receptor inhibitor protein-4.

XX KW KDEL receptor inhibitor; heat shock protein; immune response;

KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;

KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;

XX infectious disease; allergy; autoimmune disease.

OS Chimeric - Mus sp.

OS Chimeric - Xenopus sp.

OS Chimeric - Camelus sp.

OS Chimeric - Rattus sp.

XX Key Location/Qualifiers

FT Peptide 1..20

FT /label= Signal peptide

FT /note= "Derived from mouse B1p"

FT Domain 26..30

FT /note= "Altered subsequence from rat COMP"

KM		infectious disease; allergy; autoimmune disease.
XX		
OS	Chimeric - Mus sp.	
OS	Chimeric - Camelus sp.	
OS	Chimeric - Rattus sp.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..20
FT		/label= Signal peptide
FT	Domain	/note= "Derived from mouse BIP"
FT		26..30
FT	Domain	/note= "Altered subsequence from rat COMP"
FT		30..71
FT	Domain	/note= "Mouse thrombospondin 3 trimerisation domain"
FT		72..95
FT	Domain	/note= "Camel IgG linker domain"
XX		
PN	WO200006729-A1.	
PD	10-FEB-2000.	
XX		
PF	28-JUL-1999;	99NO-US17147.
XX		
PR	29-JUL-1998;	98US-0124671.
XX		
PA	(SLOK ) SLOAN KETTERING INST CANCER RES.	
PI	Rothman JE, Mayhew M, Hoe MH;	
XX		
DR	WPI, 2000-195296/17.	
DR	N-PDB; AAZ50494.	
XX		
PT	Inhibitors of the KDEL receptor which comprises an oligomerization	
PT	domain useful for promoting secretion of proteins which are normally	
PT	retained within the cell -	
XX		
PS	Disclosure, Fig 3; 87pp; English.	
XX		
CC	The patent discloses the use of KDEL receptor inhibitor to promote	
CC	secretion of proteins that are normally retained within the cell such as	
CC	heat shock proteins by inhibiting KDEL receptor-mediated return of	
CC	protein complexes to endoplasmic reticulum. This makes the secreted heat	
CC	shock proteins more accessible to the immune system and improves immune	
CC	response to a target antigen. The inhibitor protein comprises several	
CC	subunits where each subunit comprises an oligomerisation domain and has	
CC	at its carboxy terminus a region which binds to a KDEL receptor. The	
CC	target antigen may be associated with diseases including neoplasia such	
CC	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and	
CC	astrocytoma, with defective tumour suppressor genes, oncogenes,	
CC	infectious diseases, allergy or autoimmune diseases. The present	
CC	sequence is KDEL receptor inhibitor comprising regions encoding a	
CC	cleavable signal peptide; the oligomerisation domain of mouse	
CC	thrombospondin 3 (TSP3) trimerisation domain; a camel IgG linker domain	
CC	and the carboxy-terminal sequence KDEL. The subsequence GDCC is an	
CC	alteration of rat cartilage oligomeric matrix protein which provides	
CC	increased stability via disulphide bonds.	
XX		
SQ	Sequence	105 AA;
Query Match	57.8%; Score 337.5; DB 21; Length 105;	
Best Local Similarity	63.7%; Pred. No. 9e-27;	
Matches	65; Conservative 12; Mismatches 18; Indels 7; Gaps 2;	
OY	8 LLAALAVCAAKKGSSISGGCCSDLPQMRLBELQETNALQDVARDWLROQVREITFLKNT 67           :           :           :           11 LLLIGAV---RAGSSSLGGDCKCAL----VTQLTLFQQIIVELRDIDRVKEMSLIRNT 63	
OY	68 VMECDACGPPOPPOPOPPOPKPPKPPEEGTSSEKDEL 109           :           :           :           Db IMECQVCGPPOPPOPKPPQPQPKPPKEPPEGTGSSEKDEL 105	
RESULT	10	

ID	AA4964
AC	AA4964; standard; Protein: 90 AA.
XX	
XX	AA4964;
DT	23-MAY-2000 (first entry)
XX	
DE	KDEL receptor inhibitor protein-7.
XX	
KW	KDEL receptor inhibitor; heat shock protein; immune response; oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia; melanoma; carcinoma; glioblastoma; astrocytoma; oncogene; infectious disease; allergy; autoimmune disease.
OS	Chimeric - Adenovirus E3. Chimeric - Homo sapiens. Chimeric - Camelus sp.
XX	
FT	Key Location/Qualifiers
FT	Peptide 1..20
FT	/note= "Derived from adenovirus E3"
FT	Domain 30..56
FT	/note= "Human phospholamban pentamerisation domain"
FT	Domain 57..80
FT	/note= "Camel Igc linker domain"
XX	
PN	WO200006729-A1.
PD	10-FEB-2000.
XX	
PF	28-JUL-1999; 99MO-US17147.
XX	
PR	29-JUL-1998; 98US-0124671.
PA	(SLOK ) SLOAN KETTERING INST CANCER RES.
XX	
PI	Rochman JE, Mayhew M, Hoe MH;
XX	
DR	WI; 2000-195296/17.
DR	N-PSDB; AAZ50498.
XX	
PT	Inhibitors of the KDEL receptor which comprises an oligomerization domain useful for promoting secretion of proteins which are normally retained within the cell -
XX	
PS	Disclosure; Fig 7; 87pp; English.
CC	The patent discloses the use of KDEL receptor inhibitor to promote secretion of proteins that are normally retained within the cell such as heat shock proteins by inhibiting KDEL receptor-mediated return of protein complexes to endoplasmic reticulum. This makes the secreted heat shock proteins more accessible to the immune system and improves immune response to a target antigen. The inhibitor protein comprises several subunits where each subunit comprises an oligomerisation domain and has at its carboxy terminus a region which binds to a KDEL receptor. The target antigen may be associated with diseases including neoplasia such as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and astrocytoma, with defective tumour suppressor genes, oncogenes, infectious diseases, allergy or autoimmune diseases. The present sequence is KDEL receptor inhibitor comprising regions encoding a cleavable signal peptide; the oligomerisation domain of human phospholamban (PLB) pentamerisation domain; a camel IgG linker domain and the carboxy-terminal sequence KDEL. The substance CDCC is an alteration of rat cartilage oligomeric matrix protein which provides increased stability via disulphide bonds.
XX	
SO	Sequence 90 AA:
Query Match	53.0%; Score 309.5; DB 21; Length 90;
Best Local Similarity	59.6%; Pred. No. 5,3e-24;
Matches 68; Conservative 1; Mismatches 16; Indels 29; Gaps	

Db 1 MRYVITLGLTAAVCSAAKKSSSLGGCCCLT-----QLRF 36

Qy 61 ITPLKNTVMECDAC-----GPQPKPQPKPQPKPQPKPEPCTGSSKDEL 109

Db 37 INFCLILICLLILICILVWLPLPQPKPQPKPQPKPQPKPEPCTGSSKDEL 90

RESULT 11

AA000040 ID AAB00040 standard; Protein; 242 AA.

AC AAB00040;

XX 08-NOV-2000 (first entry)

XX Human COMP/TSP-1 chimeric protein.

DE TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;

XX thrombospondin; angiogenesis; tumour; treatment; cancer;

KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;

KV glaucoma.

XX Homo sapiens.

OS Synthetic.

XX WO200044908-A2.

XX 03-AUG-2000.

XX 01-FEB-2000; 2000WO-US02482.

XX 01-FEB-1999; 99US-0118053.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA Lawler JW;

XX WPI; 2000-514823/46.

DR N-PSDB; AAA47734.

XX Nucleic acids encoding chimeric proteins such as cartilage oligomeric

PT matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for

PT inhibiting angiogenesis and treating diseases such as cancer

XX Claim 33; Fig 4a-b; 40pp; English.

XX New nucleic acids are described which encode a protein comprising

CC the second and third type 1 repeats of human TSP (thrombospondin)-1,

CC but not the TGF (transforming growth factor)-beta activation region

CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing

CC the second and third type-1 repeats and the COMP (cartilage

CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was

CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1

CC caused inhibition of the growth of tumours in mice models.

CC Thus the nucleic acids and proteins may be useful for treating

CC angiogenesis related diseases such as cancer (by reducing the rate of

CC growth and size of tumours), arthritis, psoriasis, diabetic

CC retinopathy, corneal graft rejection, and glaucoma. They may also be

CC used for treating human immunodeficiency virus (HIV) infection.

CC Anti-angiogenic therapy has little toxicity, does not require the

CC therapeutic agent to enter tumour cells or cross the blood-brain

CC barrier, controls tumour growth independently of growth of

CC tumour cell heterogeneity, and does not induce drug resistance.

XX Sequence 242 AA;

SO Query Match 44.9%; Score 262.5; DB 21; Length 242;

Best Local Similarity 69.5%; Pred. No. 1e-18;

Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

8 LLAIAVCSAAKKSSSLGGCCCLDPLGPMLELOETNAALDVRDWLRFQVREITFLKNT 67

Db 10 LITLALGASGGGSPLG---SDIGPOMRELQETNALQDVRDMLRQVREITFLKNT 65

QY 68 VMECDACGPQ-----POPKP 82  
 |||||  
 |:  
 Db 66 VMECDACGMQGSVRTGLPSVRP 87

RESULT 12  
 AAB00041 ID AAB00041 standard; Protein; 300 AA.  
 XX AAB00041;  
 AC  
 XX 08-NOV-2000 (first entry)  
 DT  
 XX  
 DE Human COMP/TSP-2 chimeric protein.  
 XX  
 XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;  
 KM thrombospondin; angiogenesis; tumour; treatment; cancer;  
 KM arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;  
 KM glaucoma.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX WO200044908-A2.  
 PN  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 01-FEB-2000; 2000MO-US02482.  
 XX  
 PR 01-FEB-1999; 99US-0118053.  
 XX  
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX  
 PI Lawler JW;  
 XX  
 DR WPI; 2000-514823/46.  
 DR N-PSDB; AAA47735.  
 XX  
 PT Nucleic acids encoding chimeric proteins such as cartilage oligomeric  
 matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for  
 PT inhibiting angiogenesis and treating diseases such as cancer  
 XX  
 XX Claim 50; Fig 5a-b; 40pp; English.  
 PS  
 XX  
 CC New nucleic acids are described which encode a protein comprising  
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,  
 CC but not the TGF (transforming growth factor)-beta activation region  
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing  
 CC the second and third type-1 repeats and the COMP (cartilage  
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was  
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1  
 CC caused inhibition of the growth of tumours in mice models.  
 CC Thus the nucleic acids and proteins may be useful for treating  
 CC angiogenesis related diseases such as cancer (by reducing the rate of  
 CC growth and size of tumours), arthritis, psoriasis, diabetic  
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be  
 CC used for treating human immunodeficiency virus (HIV) infection.  
 CC Anti-angiogenic therapy has little toxicity, does not require the  
 CC therapeutic agent to enter tumour cells or cross the blood-brain  
 CC barrier, controls tumour growth independently of growth of  
 CC tumour cell heterogeneity, and does not induce drug resistance.  
 CC  
 CC Sequence 300 AA:  
 SQ

Query Match 44.9%; Score 262.5; DB 21; Length 300;  
 Best Local Similarity 69.5%; Pred. No. 1,4e-18;  
 Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

QY 8 LITLALGASAAKKSISGDCSDIGPOMRELQETNALQDVRDMLRQVREITFLKNT 67  
 |||||  
 |:  
 Db 10 LITLALGASGGGSPLG---SDIGPOMRELQETNALQDVRDMLRQVREITFLKNT 65

QY 68 VMECDACGPQ-----POPKP 82  
 |||||  
 |:  
 Db 66 VMECDACGMQGSVRTGLPSVRP 87

RESULT 13  
 AAB00044 ID AAB00044 standard; Protein; 757 AA.  
 XX AAB00044;  
 AC  
 XX 08-NOV-2000 (first entry)  
 DT  
 XX  
 DE Human cartilage oligomeric matrix protein (COMP).  
 XX  
 XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;  
 KM thrombospondin; angiogenesis; tumour; treatment; cancer;  
 KM arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;  
 KM glaucoma.  
 XX  
 OS Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Region 89..128  
 FT /label= Type 2 repeat region  
 FT 129..181  
 FT Region /label= Type 2 repeat region  
 FT 182..226  
 FT /label= Type 2 repeat region  
 FT 227..268  
 FT Region /label= Type 2 repeat region  
 XX  
 PN WO200044908-A2.  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 01-FEB-2000; 2000MO-US02482.  
 XX  
 PR 01-FEB-1999; 99US-0118053.  
 XX  
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX  
 PI Lawler JW;  
 XX  
 DR WPI; 2000-514823/46.  
 XX  
 XX  
 PT Nucleic acids encoding chimeric proteins such as cartilage oligomeric  
 matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for  
 PT inhibiting angiogenesis and treating diseases such as cancer  
 XX  
 XX Disclosure; Fig 3; 40pp; English.  
 PS  
 XX  
 CC New nucleic acids are described which encode a protein comprising  
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,  
 CC but not the TGF (transforming growth factor)-beta activation region  
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing  
 CC the second and third type-1 repeats and the COMP (cartilage  
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was  
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1  
 CC caused inhibition of the growth of tumours in mice models.  
 CC Thus the nucleic acids and proteins may be useful for treating  
 CC angiogenesis related diseases such as cancer (by reducing the rate of  
 CC growth and size of tumours), arthritis, psoriasis, diabetic  
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be  
 CC used for treating human immunodeficiency virus (HIV) infection.  
 CC Anti-angiogenic therapy has little toxicity, does not require the  
 CC therapeutic agent to enter tumour cells or cross the blood-brain  
 CC barrier, controls tumour growth independently of growth of  
 CC tumour cell heterogeneity, and does not induce drug resistance.  
 CC  
 CC Sequence 757 AA:  
 SQ



Query Match 44.9%; Score 262.5; DB 21; Length 757;  
Best Local Similarity 69.5%; Pred. No. 4e-18;  
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

QY 8 LIALAAVCSAAKKSSSLGGCCSDLGPMRLRELOETNAALQDVWDMLRQVREITFLKNT 67  
DB 10 LTLTAAAGASGGGSPG---SDLGPMRLRELOETNAALQDVWDMLRQVREITFLKNT 65  
QY 68 VMECDACGPGQ-----PQPKP 82  
DB 66 VMECDACGMOOSVRTGLPSVRP 87

RESULT 14  
ABJ05594  
ID ABJ05594 standard; Protein; 757 AA.  
AC ABJ05594;  
DT 14-NOV-2002 (first entry)  
XX Breast cancer-associated protein 59.  
XX  
XX Breast cancer; breast cancer-associated gene sequence;  
XX drug development; pharmacogenetics; biosensor development.  
XX Unidentified.  
XX  
XX NO200259377-A2.  
XX  
XX 01-AUG-2002.  
XX  
XX 24-JAN-2002; 2002WO-US02242.  
XX  
XX 24-JAN-2001; 2001US-263965P.  
XX 02-FEB-2001; 2001US-265928P.  
XX 09-APR-2001; 2001US-282698P.  
XX 09-APR-2001; 2001US-0829472.  
XX 04-MAY-2001; 2001US-288590P.  
XX 29-MAY-2001; 2001US-294443P.  
XX  
XX (EOSB-) EOS BIOTECHNOLOGY INC.  
XX  
XX Mack DH, Gish KC, Afar D;  
XX WPI: 2002-583738/62.  
XX N-PSDB; ABT07751.  
XX  
XX Detecting a breast cancer-associated transcript in a patient's cell,  
XX useful for diagnosing breast cancer, comprises contacting a biological  
XX sample with a polynucleotide that selectively hybridizes with breast  
XX cancer nucleic acids -  
XX  
XX Disclosure; Page 401; 414pp; English.  
XX  
XX The invention comprises a method of detecting a breast cancer-associated  
XX transcript in a cell from a patient. The method of the invention involves  
XX contacting a biological sample from the patient with a nucleotide that  
XX hybridizes to one of the 69 breast cancer-associated gene sequences shown  
XX in the specification. The method of the invention is useful in the  
XX diagnosis or prognosis of breast cancer, and for detecting genes that are  
XX up or down-regulated in breast cancer cells. Genes identified by the  
XX method of the invention can be used in diagnostic purposes and also as  
XX targets for screening for therapeutic compounds that modulate breast  
XX cancer (e.g. hormones or antibodies). Identification of genes that are  
XX over or under expressed in breast cancer can additionally provide high-  
XX resolution, high-sensitivity datasets which can be used in the areas of  
XX diagnostics, therapeutics, drug development, pharmacogenetics, protein  
XX structure and biosensor development. Amino acid sequences ABJ05536 -  
XX ABJ05604 represent the proteins encoded by the 69 breast cancer-  
XX associated genes of the invention.  
XX  
XX Sequence 757 AA;  
XX  
XX

Query Match 44.9%; Score 262.5; DB 23; Length 757;  
Best Local Similarity 69.5%; Pred. No. 4e-18;  
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

QY 8 LIALAAVCSAAKKSSSLGGCCSDLGPMRLRELOETNAALQDVWDMLRQVREITFLKNT 67  
DB 10 LTLTAAAGASGGGSPG---SDLGPMRLRELOETNAALQDVWDMLRQVREITFLKNT 65  
QY 68 VMECDACGPGQ-----PQPKP 82  
DB 66 VMECDACGMOOSVRTGLPSVRP 87

RESULT 15  
ABR47420  
ID ABR47420 standard; Protein; 757 AA.  
AC ABR47420;  
DT 12-JUN-2003 (first entry)  
XX Breast cancer associated protein sequence SEQ ID NO:72.  
XX  
XX Breast cancer; breast cancer; cytostatic; gene therapy.  
XX Human; breast cancer; cytostatic; gene therapy.  
XX Homo sapiens.  
XX  
XX MO2003004989-A2.  
XX  
XX 16-JAN-2003.  
XX  
XX 21-JUN-2002; 2002WO-US19669.  
XX  
XX 21-JUN-2001; 2001US-299887P.  
XX 27-JUN-2001; 2001US-301572P.  
XX 18-JUL-2001; 2001US-306501P.  
XX 25-SEP-2001; 2001US-325002P.  
XX 05-MAR-2002; 2002US-362585P.  
XX 14-MAY-2002; 2002US-380391P.  
XX  
XX (MILL-) MILLENIUM PHARM INC.  
XX  
XX Lillie J, Gamavaranu M, Glatc K, Hoerth S, Kamatkar S, Mertens M;  
XX Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE, Bast RC;  
XX Hortodagyl GN, Pusztai L, Meric F, Sahin A, Mills GB;  
XX WPI: 2003-210381/20.  
XX N-PSDB; ACC50112.  
XX  
XX Breast cancer diagnosis or treatment by comparing the level of  
XX expression of a marker in a patient sample with that in the control  
XX non-breast cancer sample -  
XX  
XX Claim 1; SEQ ID 72; 128pp; English.  
XX  
XX The present invention describes a method for assessing whether a patient  
XX is afflicted with breast cancer. The method comprises comparing the level  
XX of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and  
XX ABR47386 to ABR47632) in a patient sample and the normal level of  
XX expression of the marker in a control non-breast cancer sample, where a  
XX significant increase in the level of expression of the marker in the  
XX patient sample and the normal level is an indication that the patient is  
XX afflicted with breast cancer. The breast cancer associated sequences  
XX from the present invention have cytostatic activities and can be used in  
XX gene therapy. The method is useful for diagnosing and treating breast  
XX cancer.  
XX N.B. The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
XX Sequence 757 AA;  
XX  
XX

Query Match 44.9%; Score 262.5; DB 24; Length 757;  
Best Local Similarity 69.5%; Pred. No. 4e-18;  
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;  
OY 8 LLALAAGCSAAKGGSSLGDDCCSDLGPMRLRELOETNAALQDVDPDLRQGVREITFLKNT 67  
|||: : : |||  
Db 10 LITLALAGASGGGSPFG---SDLGPMRLRELOETNAALQDVDPDLRQGVREITFLKNT 65  
|||: : : |||  
OY 68 VMECDACGPO-----PQPKP 82  
|||: : : |||  
Db 66 VMECDACGMQGSVRTGLPSVRP 87  
|||: : : |||

Search completed: August 26, 2003, 16:04:22  
Job time : 118.717 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:05:48 ; Search time 38.5841 Seconds  
(without alignments)  
119.528 Million cell updates/sec

Title: US-09-696-872-23

Perfect score: 584  
Sequence: 1 MRWITGLLALAAVCSAKK.....PKQPKPEECTGSSKDEL 109

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/6C COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/6D COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	584	100.0	109	US-09-124-671-23	Sequence 23, Appl
2	464	79.5	109	US-09-124-671-29	Sequence 29, Appl
3	434	74.3	109	US-09-124-671-27	Sequence 27, Appl
4	428.5	72.4	115	US-09-124-671-15	Sequence 15, Appl
5	422	72.3	134	US-09-124-671-34	Sequence 34, Appl
6	414.5	71.0	115	US-09-124-671-13	Sequence 13, Appl
7	371.5	63.6	109	US-09-124-671-21	Sequence 21, Appl
8	345.5	59.2	109	US-09-124-671-19	Sequence 19, Appl
9	337.5	57.8	105	US-09-124-671-17	Sequence 17, Appl
10	309.5	53.0	90	US-09-124-671-25	Sequence 25, Appl
11	241	41.3	46	US-09-124-671-2	Sequence 2, Appl
12	234.5	40.2	66	US-09-091-814-46	Sequence 46, Appl
13	228	39.0	64	US-09-091-814-45	Sequence 45, Appl
14	211	36.1	46	US-09-124-671-1	Sequence 1, Appl
15	146	25.0	35	US-08-471-780C-38	Sequence 38, Appl
16	146	25.0	35	US-08-467-282B-38	Sequence 38, Appl
17	146	25.0	35	US-08-471-282A-38	Sequence 38, Appl
18	146	25.0	35	US-08-466-710C-38	Sequence 38, Appl
19	146	25.0	35	US-08-468-739C-38	Sequence 38, Appl
20	146	25.0	35	US-08-471-780C-44	Sequence 44, Appl
21	146	25.0	54	US-08-467-282B-44	Sequence 44, Appl
22	146	25.0	54	US-08-471-282A-44	Sequence 44, Appl
23	146	25.0	54	US-08-466-710C-44	Sequence 44, Appl
24	146	25.0	54	US-08-468-739C-44	Sequence 44, Appl
25	146	25.0	60	US-08-471-780C-87	Sequence 87, Appl
26	146	25.0	60	US-08-467-282B-87	Sequence 87, Appl
27	146	25.0	60	US-08-471-282A-87	Sequence 87, Appl

28	146	25.0	60	US-08-466-710C-87	Sequence 87, Appl
29	146	25.0	60	US-08-468-739C-87	Sequence 87, Appl
30	145.5	24.9	961	PCT-US93-11725-4	Sequence 4, Appl
31	144	24.7	24	US-09-379-297-5	Sequence 5, Appl
32	138.5	23.7	889	PCT-US93-11725-2	Sequence 2, Appl
33	121	20.7	46	US-09-124-671-5	Sequence 5, Appl
34	119	20.4	667	US-08-718-661-2	Sequence 2, Appl
35	117	20.0	46	US-09-124-671-6	Sequence 6, Appl
36	116	19.9	261	US-09-602-565-34	Sequence 34, Appl
37	112.5	19.3	907	US-08-989-299-12	Sequence 12, Appl
38	111	19.0	221	US-09-069-023-22	Sequence 22, Appl
39	109	18.7	558	US-09-252-991A-17202	Sequence 17202, A
40	108	18.5	304	US-09-252-991A-23116	Sequence 23116, A
41	107	18.3	613	US-09-345-473B-39	Sequence 39, Appl
42	102	17.5	8991	US-08-714-741-32	Sequence 32, Appl
43	101	17.3	360	US-09-252-991A-19088	Sequence 19088, A
44	99	17.0	694	US-09-328-352-7276	Sequence 7276, Ap
45	98	16.8	619	US-08-465-746-2	Sequence 2, Appl

## ALIGNMENTS

```

RESULT 1
US-09-124-671-23
; Sequence 23, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric human COMP-KDEL
US-09-124-671-23

Query Match      100.0%; Score 584; DB 3; Length 109;
Best Local Similarity 100.0%; Pred. No. 2.2e-57;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 MRWITGLLALAAVCSAKKSSIGDCCSDLPQMRLEQETNALQDVRLQGVRE 60
Db      1 MRWITGLLALAAVCSAKKSSIGDCCSDLPQMRLEQETNALQDVRLQGVRE 60

Cy      61 ITFLKNTVMCDACGPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 109
Db      61 ITFLKNTVMCDACGPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQPPQ 109

RESULT 2
US-09-124-671-29
; Sequence 29, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29

```

```

; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric human TSP4-KDEL
US-09-124-671-29

Query Match
Best Local Similarity 79.5%; Score 464; DB 3; Length 109;
Matches 85; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

QY 1 MRYMILGLIALAAVCSAAKGGSLGGCCSDLGPMRLRELOETNALQDVRLMRAQVRE 60
DB 1 MRYMILGLIALAAVCSAAKGGSLGGCCSDGDFNRQFLGQMTQNLQDLGEVKDLRQVKE 60

QY 61 ITFLKNTWECDCGPGPQPKPQPKPQPKPQPKPEPEGTSSSEKDEL 109
DB 61 TSFLKNTTAECCGACGPGPQPKPQPKPQPKPQPKPEPEGTSSSEKDEL 109

RESULT 3
US-09-124-671-27
; Sequence 27, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric human TSP3-KDEL
US-09-124-671-27

Query Match
Best Local Similarity 74.3%; Score 434; DB 3; Length 109;
Matches 79; Conservative 13; Mismatches 17; Indels 0; Gaps 0;

QY 1 MRYMILGLIALAAVCSAAKGGSLGGCCSDLGPMRLRELOETNALQDVRLMRAQVRE 60
DB 1 MRYMILGLIALAAVCSAAKGGSLGGCCGEGTKALVTQLTFNQILVELRDIRDYKE 60

QY 61 ITFLKNTWECDCGPGPQPKPQPKPQPKPQPKPEPEGTSSSEKDEL 109
DB 61 MSLIRNTMECVCGPGPQPKPQPKPQPKPQPKPEPEGTSSSEKDEL 109

RESULT 4
US-09-124-671-15
; Sequence 15, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

```

; FEATURE:
; OTHER INFORMATION: chimeric rat COMP-KDEL
US-09-124-671-15

Query Match
Best Local Similarity 73.4%; Score 428.5; DB 3; Length 115;
Matches 87; Conservative 3; Mismatches 5; Indels 17; Gaps 3;

QY 8 LLALAAVCSAAKGGSLGGCCSDLGPMRLRELOETNALQDVRLRQVREITFLKNT 67
DB 11 LLILGAV---RAEGSSLGDDCC---PQMLRELOETNALQDVRELRQVKEITFLKNT 63

QY 68 VMECDACG-----PQPKPQPKPQPKPQPKPQPKPEPEGTSSSEKDEL 109
DB 64 VMECDACGMQPARTPGTSPPQPKPQPKPQPKPQPKPEPEGTSSSEKDEL 115

RESULT 5
US-09-124-671-34
; Sequence 34, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: KDEL/myc
US-09-124-671-34

Query Match
Best Local Similarity 72.3%; Score 422; DB 3; Length 134;
Matches 82; Conservative 2; Mismatches 1; Indels 14; Gaps 2;

QY 21 GSSLGDDCCSDLGPMRLRELOETNALQDVRLRQVREITFLKNTVMECDACG----- 75
DB 40 GSSLGDDCC---PQMLRELOETNALQDVRELRQVKEITFLKNTVMECDACGMQPAR 95

QY 76 -----PQPKPQPKPQPKPQPKPQPKPEPEGTSSSEKDEL 109
DB 96 TPGTSPQPKPQPKPQPKPQPKPQPKPEPEGTSSSEKDEL 134

RESULT 6
US-09-124-671-13
; Sequence 13, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric rat comp
US-09-124-671-13

```

```

Query Match      53.0%  Score 309.5; DB 3; Length 90;
Best Local Similarity 59.6%  Pred. No. 4.1e-27;
Matches 68; Conservative 1; Mismatches 29; Gaps 2.

```





This Page Blank (unspc'd)



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:02:18 ; Search time 38.5841 Seconds  
(without alignments)  
372.792 Million cell updates/sec

Title: US-09-696-872-23

Perfect score: 584

Sequence: 1 MRWIIIGLAAVCSAAK.....PKPQKPEPCTGSSEKDEL 109

Scoring table: BLASTUM62  
Gapop 10.0 , Gapext 0.5

Searched: 497079 seqs, 131961718 residues

Total number of hits satisfying chosen parameters: 497079

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262.5	44.9	242	10	US-09-919-603-5
2	262.5	44.9	300	10	US-09-919-603-7
3	262.5	44.9	757	10	US-09-919-603-3
4	262.5	44.9	757	12	US-10-301-822-41
5	262.5	44.9	757	15	US-10-177-293-72
6	145.5	24.9	961	15	US-10-017-721-4
7	140	24.0	178	15	US-10-154-971-22
8	139	23.8	303	15	US-10-154-971-24
9	126	21.6	578	15	US-10-156-761-12544
10	122	20.9	400	12	US-10-301-822-155
11	116	19.9	261	15	US-10-280-953-17
12	115	19.7	285	10	US-09-881-752A-228
13	111	19.0	807	14	US-10-138-221-7
14	108	18.5	316	10	US-09-764-864-967
15	107.5	18.4	266	10	US-09-738-626-6457

16	107	18.3	613	10	US-09-862-027-39	Sequence 39, Appl
17	105	18.0	265	14	US-10-073-256-78	Sequence 78, Appl
18	102	17.5	417	12	US-09-949-029-2	Sequence 2, Appl
19	102	17.5	1333	9	US-09-815-242-10936	Sequence 10936, A
20	99	17.0	54	10	US-09-855-754-22	Sequence 22, Appl
21	99	17.0	827	15	US-10-171-384-3	Sequence 3, Appl
22	99	17.0	1493	9	US-09-858-754-4	Sequence 4, Appl
23	98	16.8	619	11	US-09-882-774-1	Sequence 1, Appl
24	97	16.6	250	10	US-09-252-150-20	Sequence 20, Appl
25	97	16.6	583	15	US-10-156-761-12356	Sequence 12356, A
26	95.5	16.4	605	10	US-09-311-631A-25	Sequence 25, Appl
27	94.5	16.2	183	10	US-09-252-150-2	Sequence 2, Appl
28	94	16.1	59	10	US-09-855-754-16	Sequence 16, Appl
29	94	16.1	762	11	US-09-917-378-1	Sequence 1, Appl
30	94	16.1	824	15	US-10-171-384-1	Sequence 1, Appl
31	94	16.1	840	11	US-09-884-465A-7	Sequence 7, Appl
32	94	16.1	999	11	US-09-884-465A-376	Sequence 376, Appl
33	94	16.1	999	11	US-09-884-465A-377	Sequence 377, Appl
34	94	16.1	1126	11	US-09-884-465A-383	Sequence 383, Appl
35	94	16.1	1238	11	US-09-884-465A-381	Sequence 381, Appl
36	94	16.1	1365	11	US-09-884-465A-382	Sequence 382, Appl
37	92	15.8	572	15	US-10-205-823-225	Sequence 225, Appl
38	91	15.6	52	10	US-09-855-754-17	Sequence 17, Appl
39	91	15.6	56	10	US-09-855-754-18	Sequence 18, Appl
40	91	15.6	439	11	US-09-056-019-37	Sequence 37, Appl
41	90	15.4	52	10	US-09-855-754-21	Sequence 21, Appl
42	90	15.4	324	9	US-09-815-242-11345	Sequence 11345, A
43	89.5	15.3	535	11	US-09-927-827-50	Sequence 50, Appl
44	89	15.2	87	15	US-10-215-432-27	Sequence 27, Appl
45	89	15.2	204	10	US-09-252-150-3	Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-09-919-603-5  
Sequence 5, Application US/09919603  
Patent No. US20020137679A1  
GENERAL INFORMATION:  
APPLICANT: Lawler, John W.  
TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP  
FILE REFERENCE: 1440.1033-007  
CURRENT APPLICATION NUMBER: US/09/919,603  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: PCT/US00/02482  
PRIOR FILING DATE: 2000-02-01  
PRIOR APPLICATION NUMBER: 60/118,053  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 242  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chimeric protein  
US-09-919-603-5

Query Match 44.9%; Score 262.5; DB 10; Length 242;  
Best Local Similarity 69.5%; Pred. No. 3.6e-18;  
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;  
QY 8 LLAALAVCSAAKSSIGGDCSDLGQMRLEIQTNAALQDVWDLRQVREITFLKNT 67  
DB 10 LTLAALGASGGGSPG---SDLGQMRLEIQTNAALQDVWDLRQVREITFLKNT 65  
QY 68 VMECDACGPO-----PKPKP 82  
DB 66 VMECDACGMOOSVRTGLPSVP 87

## RESULT 2

US-09-919-603-7  
; Sequence 7, Application US/09919603  
; Patent No. US20020137679A1  
; GENERAL INFORMATION:  
; APPLICANT: Lawler, John W.  
; TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP  
; FILE REFERENCE: 1440.1033-007  
; CURRENT APPLICATION NUMBER: US/09/919,603  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: PCT/US00/02482  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/118,053  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 300  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chimeric protein  
US-09-919-603-7

Query Match 44.9%; Score 262.5; DB 10; Length 300;  
Best Local Similarity 69.5%; Pred. No. 4,6e-18;  
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

QY 8 LIALAAVCSAAKKGSSLGDCSDIGPQMLRELQETNALQDVVDMLRQVREITFLKNT 67

DB 10 LITLALGASGGQSPGLG---SDLGPMRLRELQETNALQDVVDMLRQVREITFLKNT 65

QY 68 VMECDACGPO-----POPKP 82  
DB 66 VMECDACGMOQSVRTGLPSVRP 87

## RESULT 3

US-09-919-603-3  
; Sequence 3, Application US/09919603  
; Patent No. US20020137679A1  
; GENERAL INFORMATION:  
; APPLICANT: Lawler, John W.  
; TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP  
; FILE REFERENCE: 1440.1033-007  
; CURRENT APPLICATION NUMBER: US/09/919,603  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: PCT/US00/02482  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/118,053  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 757  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-919-603-3

Query Match 44.9%; Score 262.5; DB 10; Length 757;  
Best Local Similarity 69.5%; Pred. No. 1.3e-17;  
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

QY 8 LIALAAVCSAAKKGSSLGDCSDIGPQMLRELQETNALQDVVDMLRQVREITFLKNT 67

DB 10 LITLALGASGGQSPGLG---SDLGPMRLRELQETNALQDVVDMLRQVREITFLKNT 65

QY 68 VMECDACGPO-----POPKP 82  
DB 66 VMECDACGMOQSVRTGLPSVRP 87

## RESULT 4

US-10-301-822-41  
; Sequence 41, Application US/10301822  
; Publication No. US20030148410A1  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc.  
; APPLICANT: Berger, Allison  
; APPLICANT: Guillemette, Tracy L.  
; APPLICANT: Kametkar, Shubhangi  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Thibodeau, Stephen N.  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; FILE REFERENCE: MEM01-029P2RM  
; CURRENT APPLICATION NUMBER: US/10/301,822  
; PRIOR FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 60/339,971  
; PRIOR FILING DATE: 2001-12-10  
; PRIOR APPLICATION NUMBER: US 60/361,978  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/381,988  
; NUMBER OF SEQ ID NOS: 228  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 41  
; LENGTH: 757  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-301-822-41

Query Match 44.9%; Score 262.5; DB 12; Length 757;  
Best Local Similarity 69.5%; Pred. No. 1.3e-17;  
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

QY 8 LIALAAVCSAAKKGSSLGDCSDIGPQMLRELQETNALQDVVDMLRQVREITFLKNT 67

DB 10 LITLALGASGGQSPGLG---SDLGPMRLRELQETNALQDVVDMLRQVREITFLKNT 65

QY 68 VMECDACGPO-----POPKP 82  
DB 66 VMECDACGMOQSVRTGLPSVRP 87

## RESULT 5

US-10-177-293-72  
; Sequence 72, Application US/10177293  
; Publication No. US20030124128A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Glatt, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Ganawairpu, Manjula  
; APPLICANT: Kametkar, Shubhangi  
; APPLICANT: Mertens, Maureen  
; APPLICANT: Myer, Vic  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Hoersch, Sebastian  
; APPLICANT: Monahan, John  
; APPLICANT: Meyers, Rachel E.  
; APPLICANT: Baer Jr., Robert C.  
; APPLICANT: Hortobagyi, Gabriel N.  
; APPLICANT: Puzetal, Lajos  
; APPLICANT: Meric, Funda  
; APPLICANT: Sahin, Ayesgul  
; APPLICANT: Mills, Gordon B.  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
; FILE REFERENCE: MRI-038  
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER

US-10-177-293-72

Query Match 44.9%; Score 262.5; DB 15; Length 757;  
Best Local Similarity 69.5%; Pred. No. 1.3e-17;  
Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

ORGANISM: Homo sapiens

SEQUENCE: FASTSEQ for Windows Version 4.0

SEQUENCE ID NO 72

LENGTH: 757

TYPE: PRT

QY 8 L1L1AAVCSAAKSSLSGDCSDLPQMLRELOETNAALODVDMLRQVREITFLKNT 67  
DB 10 L1L1AAVCSAAKSSLSGDCSDLPQMLRELOETNAALODVDMLRQVREITFLKNT 65

QY 68 VMECDACGPQ-----PQPKP 82  
DB 66 VMECDACGMQSVRTGTPSVRP 87

RESULT 6

US-10-017-721-4

Sequence 4, Application US/10017721  
Publication No. US20030096248A1

GENERAL INFORMATION:

APPLICANT: McCarthy, Jeanette

APPLICANT: Daley, George

APPLICANT: Bolk, Stacey

TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE

FILE REFERENCE: WMI-003

CURRENT APPLICATION NUMBER: US/10/017,721

CURRENT FILING DATE: 2001-12-14

PRIOR APPLICATION NUMBER: US 60/317,033

PRIOR FILING DATE: 2001-09-04

PRIOR APPLICATION NUMBER: US 60/330,248

PRIOR FILING DATE: 2001-10-17

NUMBER OF SEQ ID NOS: 13

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4

LENGTH: 961

TYPE: PRT

ORGANISM: Homo sapiens

US-10-017-721-4

Query Match 24.9%; Score 145.5; DB 15; Length 961;  
Best Local Similarity 42.9%; Pred. No. 5.3e-06;  
Matches 30; Conservative 10; Mismatches 25; Indels 5; Gaps 2;

QY 31 DLGQMLRELOETNAALODVDMLRQVREITFLKNTVMECDACGP-----OPQPKPQPKP 86  
DB 218 DFNROFLGQMTOLNQLGSEVDLRLQVREITFLKNTIAECACGPKLKFQSPSTVAP 277

QY 87 OPQPKPQPKP 96  
DB 278 AP-PAPPTRP 286

RESULT 7

US-10-154-971-22

Sequence 22, Application US/10154971  
Publication No. US20030088074A1

GENERAL INFORMATION:

APPLICANT: Hamers, Raymond

Muyldermans, Serge

TITLE OF INVENTION: VARIABLE FRAGMENTS OF IMMUNOGLOBULINS -  
USE FOR THERAPEUTIC OR VETERINARY PURPOSES

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESSES:

ADDRESSEE: SPENCER & FRANK

STREET: 1100 New York Avenue, N.W., Suite 300 East

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/154,971

FILING DATE: 28-May-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/945,244

FILING DATE: <Unknown>

APPLICATION NUMBER: PCT/EP96/01725

FILING DATE: 25-APR-1996

APPLICATION NUMBER: EP 95400932.0

FILING DATE: 25-APR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Gollin, Michael A.

REGISTRATION NUMBER: 31,957

REFERENCE/DOCKET NUMBER: GUPLA 0003

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-414-4000

TELEFAX: 202-414-4040

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 178 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 22:

US-10-154-971-22

Query Match 24.0%; Score 140; DB 15; Length 178;  
Best Local Similarity 88.9%; Pred. No. 2.7e-06;  
Matches 24; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 76 PQPQPKPQPPQPPQPKPQPKPPEEGTG 102  
DB 138 PQPQPKPQPPQPPQPKPQPKPPEMECG 164

RESULT 8

US-10-154-971-24

Sequence 24, Application US/10154971  
Publication No. US20030088074A1

GENERAL INFORMATION:

APPLICANT: Hamers, Raymond

Muyldermans, Serge

TITLE OF INVENTION: VARIABLE FRAGMENTS OF IMMUNOGLOBULINS -  
USE FOR THERAPEUTIC OR VETERINARY PURPOSES

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESSES:

ADDRESSEE: SPENCER & FRANK

STREET: 1100 New York Avenue, N.W., Suite 300 East

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/154,971  
FILING DATE: 28-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/945,244  
FILING DATE: <Unknown>  
APPLICATION NUMBER: PCT/EP96/01725  
FILING DATE: 25-APR-1996  
APPLICATION NUMBER: EP 95400932.0  
FILING DATE: 25-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Gollin, Michael A.  
REGISTRATION NUMBER: 31,957  
REFERENCE/DOCKET NUMBER: GUPLA 0003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-414-4000  
TELEFAX: 202-414-4040  
INFORMATION FOR SEQ ID NO: 24:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
US-10-154-971-24

Query Match 23.8% Score 139; DB 15; Length 303;  
Best Local Similarity 100.0%; Pred. No. 6.2e-06;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 PGPQPKPQPQPQPKPQPKPEP 98  
DB 138 PGPQPKPQPQPQPKPQPKPEP 160

RESULT 9  
US-10-156-761-12544  
Sequence 12544, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 12544  
LENGTH: 578  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-12544

Query Match 21.6% Score 126; DB 15; Length 578;  
Best Local Similarity 79.2%; Pred. No. 0.00024;  
Matches 19; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 76 PGPQPKPQPQPQPKPQPKPEP 99  
DB 36 PGPQPKPQPQPQPKPQPKPEP 59

RESULT 10  
US-10-301-822-155  
Sequence 155, Application US/10301822  
Publication No. US20030148410A1  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc.  
APPLICANT: Berger, Allison  
APPLICANT: Guillemette, Tracy L.  
APPLICANT: Kamatkar, Shubhangi  
APPLICANT: Schlegel, Robert  
APPLICANT: Monahan, John E.  
APPLICANT: Thibodeau, Stephen N.  
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
THERAPY OF COLON CANCER  
FILE REFERENCE: MPW01-029P2RNM  
CURRENT APPLICATION NUMBER: US/10/301,822  
CURRENT FILING DATE: 2002-11-21  
PRIOR APPLICATION NUMBER: US 60/339,971  
PRIOR FILING DATE: 2001-12-10  
PRIOR APPLICATION NUMBER: US 60/361,978  
PRIOR FILING DATE: 2002-03-05  
PRIOR APPLICATION NUMBER: US 60/381,988  
PRIOR FILING DATE: 2002-05-20  
NUMBER OF SEQ ID NOS: 228  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 155  
LENGTH: 400  
TYPE: PRT  
ORGANISM: Homo Sapiens  
US-10-301-822-155

Query Match 20.9% Score 122; DB 12; Length 400;  
Best Local Similarity 79.2%; Pred. No. 0.0004;  
Matches 19; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 76 PGPQPKPQPQPQPKPQPKPEP 99  
DB 322 PGPQPKPQPQPQPKPQPKPEP 345

RESULT 11  
US-10-280-953-17  
Sequence 17, Application US/10280953  
Publication No. US20030113317A1  
GENERAL INFORMATION:  
APPLICANT: Yue, Henry  
APPLICANT: Arvizu, Chandra S.  
APPLICANT: Corley, Neil C.  
APPLICANT: Guegler, Karl J.  
TITLE OF INVENTION: MOLECULE ASSOCIATED WITH APOPTOSIS  
FILE REFERENCE: PC-0018-1 CIP  
CURRENT APPLICATION NUMBER: US/10/280,953  
CURRENT FILING DATE: 2002-10-23  
PRIOR APPLICATION NUMBER: 09/602,565  
PRIOR FILING DATE: 2000-06-22  
PRIOR APPLICATION NUMBER: 09/106,920  
PRIOR FILING DATE: 1998-06-29  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PERL Program  
SEQ ID NO 17  
LENGTH: 261  
TYPE: PRT  
ORGANISM: Mus musculus  
FEATURE: misc.feature  
NAME/KEY: incycle ID No. US20030113317A1 g1469400  
OTHER INFORMATION: incycle ID No. US20030113317A1 g1469400

US-10-280-953-17

Query Match 19.9%; Score 116; DB 15; Length 261;

Best Local Similarity 75.0%; Pred. No. 0.00095;

Matches 18; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 76 POPPKPOPOPKPOPKPEPE 99

Db 173 POPPOIQPOPOPOPOPOPOPO 196

RESULT 12

US-09-881-752A-228

Sequence 228, Application US/09881752A

Patent No. US20020115078A1

GENERAL INFORMATION:

APPLICANT: Kleanthous, Harold

APPLICANT: Al-Garawi, Amal

APPLICANT: Miller, Charles

APPLICANT: Tomb, Jean-Francois

APPLICANT: Oomen, Raymond P.

TITLE OF INVENTION: Identification of Polynucleotides

TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the

FILE REFERENCE: 06132/041002

CURRENT APPLICATION NUMBER: US/09/881,752A

CURRENT FILING DATE: 2001-06-15

PRIOR APPLICATION NUMBER: US 08/833,457

PRIOR FILING DATE: 1997-04-01

NUMBER OF SEQ ID NOS: 370

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 228

LENGTH: 285

TYPE: PRT

ORGANISM: Helicobacter pylori

US-09-881-752A-228

Query Match 19.7%; Score 115; DB 10; Length 285;

Best Local Similarity 66.7%; Pred. No. 0.0013;

Matches 16; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 76 POPPKPOPOPKPOPKPEPE 99

Db 106 PKPKPKPKPKPKPKPKPKPKPK 129

RESULT 13

US-10-138-221-7

Sequence 7, Application US/10138221

Publication No. US20020199216A1

GENERAL INFORMATION:

APPLICANT: Amy F. Macrae

TITLE OF INVENTION: USE OF TRANSPOSABLE ELEMENTS FOR ALTERING GENE EXPRESSION

FILE REFERENCE: 51178/7

CURRENT APPLICATION NUMBER: US/10/138,221

CURRENT FILING DATE: 2002-08-01

PRIOR APPLICATION NUMBER: US 60/287,882

PRIOR FILING DATE: 2001-05-01

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn version 3.1

SEQ ID NO 7

LENGTH: 807

TYPE: PRT

ORGANISM: Zea mays

US-10-138-221-7

Query Match 19.0%; Score 111; DB 14; Length 807;

Best Local Similarity 73.9%; Pred. No. 0.011;

Matches 17; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 77 QOPKPOPOPOPKPKPEPE 99

Db 108 EPOPOPOPOPOPOPOPEPEE 130

RESULT 14

US-09-764-864-967

Sequence 967, Application US/09764864

Patent No. US20020132753A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT23

CURRENT APPLICATION NUMBER: US/09/764,864

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 967

LENGTH: 316

TYPE: PRT

ORGANISM: Homo sapiens

US-09-764-864-967

Query Match 18.5%; Score 108; DB 10; Length 316;

Best Local Similarity 70.8%; Pred. No. 0.0072;

Matches 17; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 76 POPPKPOPOPKPKPEPE 99

Db 231 POPPOPPPOPEPRSPQPEPE 254

RESULT 15

US-09-738-626-6457

Sequence 6457, Application US/09738626

Publication No. US20020197605A1

GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIALI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAOKO

APPLICANT: SENO, AKIHIRO

APPLICANT: IKEDA, MASATO

APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738,626

CURRENT FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059

SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 6457

LENGTH: 266

TYPE: PRT

ORGANISM: Corynebacterium glutamicum

US-09-738-626-6457

Query Match 18.4%; Score 107.5; DB 10; Length 266;

Best Local Similarity 29.4%; Pred. No. 0.0067;

Matches 25; Conservative 16; Mismatches 29; Indels 15; Gaps 3;

Qy 39 ELQETNALQDV--RDW-----LRQVREITFLKXNVMECDACGPOQ-----PKQ 83

Db 138 QLEEMRTQLSELSEKREKGYEPTMLRAEARLILLESQQLSQQFOAPQBPVAVPEPM 197

Qy 84 POPPKPOPKPEPECTGSSEKD 108

Db 84 POPPKPOPKPEPECTGSSEKD 108

Db 198 PEPAPPEPPEPPEVAVAEABEE 222

Search completed: August 26, 2003, 16:07:06  
Job time : 39.5841 secs

---

213.080 Million cell updates/sec

2-23

LAAVCSAKK.....PK

Gapext 0.5

283308

00000

ch 08

at 45 summaries

er of results predicted equal to the score of analysis of the total score

DB	ID
2	A44315
2	S33558
1	TSHUP4
1	A45441
1	OZ2QWB
1	OZ2QKB
1	WMBEH6
2	T14355
2	S58232
2	T24470
2	C91067
2	T17557
2	F83634
2	F71829
2	B64687
2	S14959
2	S09372
1	TOZMCA
1	T15792
2	T09546
2	T16348
1	FAUTPC
2	A26036
2	B27863
2	S70010
2	A26918
2	A44418
2	S70009
2	B82478

probable transposase  
hypothetical protease  
hypothetical protease  
protein T3P18\_7 [11]  
conserved hypothetical  
probable chloroplast  
tonB protein (limpO)  
hypothetical protease  
probable serine/threonine  
conserved hypothetical  
Clumping factor B  
hypothetical protease  
hypothetical protease  
DNA binding protein  
hypothetical 47.8 kDa  
proline-rich protease

change 02-Aug-2002  
naturally related t  
1; PID:g297439  
2)  
h 755;  
els 4; Gaps  
DDMLROOVRREITFLKNT 67  
: ||: ||: ||: ||: ||  
EELLRRHVKETFLKNT 64  
change 11-Jan-2000  
binson, G.; Hamers,  
ains.  
ology





C:Accession: S07873; S12571  
R:Lockyer, M.J.; Davies, C.S.; Subhler, A.; Sinden, R.E.  
Nucleic Acids Res. 18: 376, 1990  
A:Title: Nucleotide sequence of the Plasmodium berghei circumsporozoite protein gene from  
A:Reference number: S07873; MUID:90221834; PMID:2183186  
A:Accession: S07873  
A:Molecule type: DNA  
A:Residues: 1-348 <LOC>  
A:Cross-references: EMBL:X17606  
R:Lockyer, M.J.  
submitted to the EMBL Data Library, November 1989  
A:Reference number: S12571  
A:Accession: S12571  
A:Molecule type: DNA  
A:Residues: 1-59, '1', 61-81, 83-348 <LOC2>  
A:Cross-references: EMBL:X17606; NID:97984; PIDN:CAA35608.1; PID:99785  
C:Superfamily: circumsporozoite protein; circumsporozoite type 1 repeat homology  
C:Keywords: tandem repeat  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-348/Product: circumsporozoite protein #status predicted <MAT>  
F:94-205/Region: 8-residue repeats  
F:215-247/Region: 2-residue repeats  
F:274-326/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match 22.8%; Score 133; DB 1; Length 348;  
Best Local Similarity 62.5%; Pred. No. 0.00017;  
Matches 20; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

Oy 76 PQPQPKPQPPQPPQPKPQPKPEPGTGSSEKD 107  
|||:|||||:|||||:|||||:|::|:  
Db 225 PQPQPPQPPQPPQPPQPPQPPQPPQGGANNKKNN 256

RESULT 7  
WMBEH6  
UL36 protein - human herpesvirus 1 (strain 17)  
C:Species: human herpesvirus 1  
C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 16-Jun-2000  
C:Accession: J10085  
R:McGeoch, D.U.; Dalrymple, M.A.; Davison, A.J.; Dolan, A.; Frame, M.C.; McNab, D.; Perrin,  
J. Gen. Virol. 69, 1511-1574, 1988  
A:Title: The complete DNA sequence of the long unique region in the genome of herpes sim  
A:Reference number: A30083; MUID:88274327; PMID:2835594  
A:Accession: J10085  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-3164 <MCG>  
C:Genetics:  
C:Cross-references: GB:X14112; NID:G1944536; PIDN:CAA32311.1; PID:G59536; GB:D00317  
A:Gene: UL36  
C:Superfamily: varicella-zoster virus gene 22 protein

Query Match 22.6%; Score 132; DB 1; Length 3164;  
Best Local Similarity 66.7%; Pred. No. 0.0017;  
Matches 22; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Oy 67 TVMECDACGPPQPPQPPQPPQPPQPPQPPQPPQ 99  
|||:|||||:|||||:|||||:|||||:  
Db 2904 TVSRSLAPQPPQPPQPPQPPQPPQPPQPPQPPQ 2936

RESULT 8  
T14355  
protein-tyrosine-phosphatase (EC 3.1.3.48) TD14 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T14355  
R:Caio, L.; Zhang, L.; Ruiz-Lozano, P.; Yang, Q.; Chien, K.R.; Graham, R.M.; Zhou, M.  
J. Biol. Chem. 273, 21077-21083, 1998  
A:Title: A novel putative protein-tyrosine phosphatase contains a BNO1-like domain and s  
A:Reference number: Z18004; MUID:98361981; PMID:9694860  
A:Accession: T14355  
A:Status: preliminary; translated from GB/EMBL/DBJ

```

A:Molecule type: mRNA
A:Residues: 1-1494 <CAO>
A:Cross-references: EMBL:AF077000; NID:g3598973; PID:g3598974; PIDN:AAC62959.1
A:Experimental source: brain
C:Genetics:
A:Gene: PTP-TD14
C:Function:
A:Description: may be involved in regulating Ha-ras-dependent cell growth
C:Keywords: phosphoric monoester hydrolase

Query Match      21.2%; Score 124; DB 2; Length 1494;
Best Local Similarity 79.2%; Pred. No. 0.0043;
Matches 19; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY      76 POPOPKPOPOPOPKPOPKPEPE 99
      |||||:||||:||||:||||:|
Db      797 POPOPPPOPOPOPOPOPOPOPO 820

RESULT 9
S58222
PO-rich protein - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 05-Nov-1999
C:Accession: S58222
R:Wagner, F.F.; Flegel, W.A.
submitted to the EMBL Data Library, July 1995
A:Description: A cDNA, which predicts a protein with PO-rich repeats, isolated from a phage
A:Reference number: S58222
A:Accession: S58222
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-400 <MAC>
A:Cross-references: EMBL:Z50194; NID:g929659; PIDN:CAA90576.1; PID:g929660

Query Match      20.9%; Score 122; DB 2; Length 400;
Best Local Similarity 79.2%; Pred. No. 0.0018;
Matches 19; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY      76 POPOPKPOPOPOPKPOPKPEPE 99
      |||||:||||:||||:||||:|
Db      322 POPOPOPOPOPOPOPOPKPOPO 345

RESULT 10
T24470
hypochemical protein T04F8.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T24470
R:Lennard, N.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19895
A:Accession: T24470
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-165 <WIL>
A:Cross-references: EMBL:Z66565; PIDN:CAA91483.1; GSPDB:GN00028; CESP:T04F8.8
A:Experimental source: clone T04F8
C:Genetics:
A:Gene: CESP:T04F8.8
A:Map position: X
A:introns: 18/2; 63/1

Query Match      20.2%; Score 118; DB 2; Length 165;
Best Local Similarity 66.7%; Pred. No. 0.0017;
Matches 16; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

OY      76 POPOPKPOPOPOPKPOPKPEPE 99
      ||||:||||:||||:||||:|
Db      108 PEPBPBPBPBPBPBPBPBPBPBP 131

```

```

RESULT 11
C91067
hypothetical protein ECe3507 [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050952)
C:/Species: Escherichia coli
C:/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:/Accession: C91067
R:/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA: Ref. 8, 11-22, 2001
A:/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic
A:/Reference number: A99629; MUID:21156231; PMID:11258796
A:/Accession: C91067
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-422 <HAY>
A:/Cross-references: GB:BA000007; PTDN:BA36930.1; PID:g13362978; GSPDB:GN00154
A:/Experimental source: strain O157:H7, substrain RIMD 050952
C:/Genetics:
C:/Gene: ECe3507

```

```

Query Match          20.2%; Score 118; DB 2; Length 422;
Best Local Similarity 53.1%; Pred. No. 0.0042;
Matches 17; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

QY      76 PQPQPQPPOPQPQPQPKEPEPTGSSEKD 107
        |||:||||::||::|||    |||:
Db       367 PEPPEPEPEPEPEPEPEPEPIRSSLKE 398

RESULT 12
T17557
procyclin homolog A67R - Chlorella virus PCV-1
C:Species: Chlorella virus PCV-1
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17557
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T17557
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-309 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96435..1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Gene: A67R

Query Match          20.0%; Score 117; DB 2; Length 309;
Best Local Similarity 34.5%; Pred. No. 0.0038;
Matches 30; Conservative 8; Mismatches 25; Indels 24; Gaps 4;

QY      21 GSSLSGGCCDGLGQMRELQETNALQDYRDWLKQVRRTFLKATVMEC--DACGPDP 78
        |||||||::|||::|:
Db       104 GSSLSGESCYTD-GWM-----QTEDIASD-----NDVYLCSMSSSPSP 141

QY      79 QPKPQPQPPOPQPQPKEPEPTGSSE 105
        :||::|||::|||::|||
Db       142 EPSPPEPSPEPSPEPSPEPTSPPE 168

RESULT 13
F83634
hypothetical protein PA0081 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83634
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Biddleman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, .J.; Loyt, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho
A:Accession: F83634
```

A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-497 <STO>  
A:/Cross-references: GB:AE004447, GB:AE004091, NID:g9945902, PIDN:AA03471.1, GSPDB:GN001:  
A:/Experimental source: strain PA01  
C:/Genetics:  
A:/Gene: PA0081

Query Match	20.0%	Score 117;	DB 2;	Length 437;
Best Local Similarity	81.8%	Pred. NO. 0.0061;		
Matches 18; Conservative	4;	Mismatches 0;	Indels 0;	Gaps 0;

```
QY      77 QPQPKPQPQPQPKPQP 98
         |||:|||||:|:|
Db      264 QPQPQPQPQPQPQP 285
```

RESULT 14  
F71829  
siderophore-mediated iron transport protein - Helicobacter pylori (strain J99)  
C/Species: Helicobacter pylori  
A/Variety: strain J99  
C/Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 21-Jan-2000  
C/Accession: F71829  
R/Alm, R.A.; Iino, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Miller, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathc  
A/Reference number: A71800; MUID:99120557; PMID:9923682  
A/Accession: F71829  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-280 <ARN>  
A/Cross-references: GB:AE001549; GB:AE001439; NID:g4155858; PIDN:AD0603.1, PID:g4155861  
A/Experimental source: strain J99  
C/Genetics:  
A/Gene: tonb\_2  
C/Superfamily: tonb protein

Query Match	19.7%	Score 115	DB 2	Length 280
Best Local Similarity	54.5%	Prod. No. 0.0052		
Matches 18	Conservative 8	Mismatches 7	Indels 0	Gaps 0

RESULT 15  
E646b7  
siderophore-mediated iron transport protein - *Helicobacter pylori* (strain 26695)  
C/Species: *Helicobacter pylori*  
C/Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #ext\_change 21-Jan-2000  
C/Accession: E646b7  
R/Tomb: J.F., White, O., Kertavagse, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khakh, H.G., Glodek, A., McKenney, S., Venter, A., Adams, J.M., Cotton, M.D., Feldman, J.M., Fujii, C., Bowman, C., Watthey, L. son, J.D., Kelley, J.M.; Cotton, M.D.; Feldman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997  
A/Authors: Wallin, E., Hayes, W.S., Borodovsky, M., Karpk, P.D., Smith, H.O., Fraser, C.N.  
A/Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.  
A/Reference number: A64520; MUID:97394467; PMID:9252185  
A/Accession: E646b7  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-285 <TOM>  
A/Cross-references: GB:AE000635; GB:AE000511; NID:g2314502; PIDN:AA006383.1; PID:g2314506  
A/Superfamily: tomB protein

Query Match	19.7%;	Score 115;	DB 2;	Length 285;
Best Local Similarity	66.7%;	Pred. No. 0.0053;		
Matches 16;	Conservative	8;	Mismatches	0;
			Indels	0;
			Gaps	0;

Wed Aug 27 10:01:26 2003

us-09-696-872-23.rpr

Page 5

D<sub>b</sub>

:   :   :   :   :	
106 PKPKPKPEPKPEPKPEPK	129

Search completed: August 26, 2003, 16:05:21  
Job time : 51.1947 secs

This page Blank (unspoiled)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:01:13 ; Search time 27.0089 Seconds  
(without alignments)  
189.786 Million cell updates/sec

Title: US-09-696-872-23  
Perfect score: 584  
Sequence: 1 MRVWILGLALAAVCSAAK.....PKPQKPEPEGTSGSEKDEL 109

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262.5	44.9	757	1 COMP_HUMAN	P49747 homo sapien
2	234	40.1	755	1 COMP_RAT	P35444 rattus norv
3	145.5	24.9	961	1 TSP4_HUMAN	P35443 homo sapien
4	144.5	24.7	980	1 TSP4_RAT	P49744 rattus norv
5	138.5	23.7	955	1 TSP4_XENLA	O06441 xenopus lae
6	133	22.8	339	1 CSP_PLABE	P06915 plasmodium
7	133	22.8	347	1 CSP_PLABA	P23093 plasmodium
8	132	22.6	3164	1 TEGU_HSV11	P10220 herpes simp
9	117	20.0	220	1 NOL3_MOUSE	O941X0 mus musculu
10	115	19.7	280	1 TONB_HELPJ	O941X0 mus musculu
11	115	19.7	285	1 TONB_HELPJ	O941X0 mus musculu
12	112.5	19.3	839	1 TRAB_MAIZE	O25899 helicobacte
13	111	19.0	115	1 PARB_TRYBB	P03010 zea mays (m
14	111	19.0	129	1 PARB_TRYBB	P14043 trypanosoma
15	111	19.0	143	1 PAR1_TRYBB	P08791 trypanosoma
16	111	19.0	145	1 PARC_TRYBB	P08469 trypanosoma
17	111	19.0	221	1 NOL3_RAT	O06084 trypanosoma
18	111	19.0	806	1 TRAI_MAIZE	O06881 rattus norv
19	108	18.5	722	1 Z219_HUMAN	P08770 zea mays (m
20	107	18.3	613	1 PKPA_PHYHL	O9P2Y4 homo sapien
21	104	17.8	1569	1 YPJA_ECOLI	O01577 phycomyces
22	102	17.5	449	1 ARG_BRANA	P40603 brassica na
23	101	17.3	342	1 TONB_PSENE	O51368 pseudomonas
24	101	17.3	450	1 V3OK_BYDVP	P09516 barley yell
25	101	17.3	565	1 TSP3_MOUSE	O70324 mus musculu
26	100.5	17.2	956	1 TSP3_MOUSE	P49746 homo sapien
27	100	17.1	643	1 VP40_HSV2	P52369 equine herp
28	100	17.1	956	1 TSP3_MOUSE	O05895 mus musculu
29	99	17.0	1493	1 M3K1_RAT	O62925 rattus norv
30	98	16.8	1520	1 ACPD_ECOLI	O48837 escherichia
31	98	16.8	534	1 APC_APATH	P46602 arabidopsis
32	97	16.6	243	1 TONB_KLEFN	P45610 klebsiella
33	96.5	16.5	737	1 CTCF_RAT	O941d1 rattus norv

## ALIGNMENTS

RESULT 1	COMP_HUMAN	STANDARD	PRT	757 AA.
ID	COMP_HUMAN	STANDARD	PRT	757 AA.
AC	P49747; O16388; O16389;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Cartilage oligomeric matrix protein precursor (COMP).			
GN	COMP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Cartilage;			
RX	MEDLINE=95229140; PubMed=7713493.			
RA	Newton G., Weremowicz S., Morton C.C., Copeland N.G.,			
RA	Gilbert D.J., Jenkins N.A., Lawler J.;			
RT	"Characterization of human and mouse cartilage oligomeric matrix			
RT	protein.";			
RL	Genomics 24:435-439(1994).			
RN	[2]			
RP	VARIANTS PSACH SER-459 DEL; TYR-468 AND TYR-472.			
RC	MEDLINE=95400301; PubMed=7670471;			
RA	Hecht J.T., Nelson L.D., Crowder E., Wang Y., Elder F.F.B.,			
RA	Harrison W.R., Francosano C.A., Prange C.K., Lennon G.G., Deere M.,			
RA	Lawler J.;			
RT	"Mutations in exon 17B of cartilage oligomeric matrix protein (COMP)			
RT	cause pseudochondroplasia.";			
RL	Nat. Genet. 10:325-329(1995).			
RN	[3]			
RP	VARIANTS MED TYR-342, AND VARIANT PSACH ARG-328.			
RC	MEDLINE=95400302; PubMed=7670472.			
RA	Briggs M.D., Hoffman S.M.G., King L.M., Olsen A.S., Mohrenweiser H.,			
RA	Leroy J.G., Mortier G.R., Rimoin D.L., Lachman R.S., Gaines E.S.,			
RA	Cerlentiak J.A., Knowlton R.G., Cohn D.H.;			
RT	"Pseudochondroplasia and multiple epiphyseal dysplasia due to			
RT	mutations in the cartilage oligomeric matrix protein gene.";			
RL	Nat. Genet. 10:330-336(1995).			
RN	[4]			
RP	VARIANT MED LYS-523.			
RC	MEDLINE=97171141; PubMed=9021009;			
RA	Ballo R., Briggs M.D., Cohn D.H., Knowlton R.G., Beighton P.H.,			
RA	Ramegar R.S.;			
RT	"Multiple epiphyseal dysplasia, ribbing type: a novel point mutation			
RT	in the COMP gene in a South African family.";			
RL	Am. J. Med. Genet. 68:396-400(1997).			
RN	[5]			
RP	VARIANT MED SER-371, AND VARIANT PSACH 513-VAL- LYS-516 DEL.			
RC	MEDLINE=97327574; PubMed=9184241;			
RA	Susic S., McGroarty J., Ahler J., Cole W.G.;			
RT	"Multiple epiphyseal dysplasia and pseudochondroplasia due to novel			
RT	mutations in the calmodulin-like repeats of cartilage oligomeric			
RT	matrix protein.";			
RL	Clin. Genet. 51:219-224(1997).			

RN [6] VARIANTS PSACH AND MED.  
 RX MEDLINE=98110533; PubMed=9463320;  
 RA Briggs M.D., Mottler G.R., Cole W.G., King L.M., Golik S.S.,  
 Bonaventura J., Nuytting L., de Paeppe A., Leroy J.G., Bleescker L.,  
 Lipson M., Wilcox W.R., Lachman R.S., Rimoin D.L., Knowlton R.G.,  
 RA Cohn D.H.;  
 RT "Diverse mutations in the gene for cartilage oligomeric matrix protein  
 RT in the pseudochondroplasia-multiple epiphyseal dysplasia disease  
 RT spectrum.";  
 RL Am. J. Hum. Genet. 62:311-319(1998).  
 RN [7]  
 RN VARIANTS PSACH AND MED.  
 RP MEDLINE=99118868; PubMed=9921895;  
 RX Ikegawa S., Ohashi H., Nishimura G., Kim K.C., Samnolte A.,  
 RA Kimizuka M., Fukushima Y., Nagai T., Nakamura Y.;  
 RA Knowlton R.G., Cohn D.H., Sykes B.;  
 RT "Identification of five novel mutations in cartilage oligomeric  
 RT matrix protein gene in pseudochondroplasia and multiple epiphyseal  
 RT dysplasia.";  
 RL Hum. Mutat. Suppl. 1:S10-S17(1998).  
 RN [9]  
 RN VARIANTS PSACH AND MED.  
 RP MEDLINE=99112405; PubMed=9452026;  
 RX Loughlin J., Ireen C., Mubareka Z., Briggs M.D., Carr A., Lynch S.-A.,  
 RA Knowlton R.G., Cohn D.H., Sykes B.;  
 RT "Identification of five novel mutations in cartilage oligomeric  
 RT matrix protein gene in pseudochondroplasia and multiple epiphyseal  
 RT dysplasia.";  
 RL Hum. Mutat. Suppl. 1:S10-S17(1998).  
 CC -1 SUBUNIT: PENTAMER; DISULFIDE-LINKED.  
 CC -1 DISEASE: DEFECTS IN COMP ARE THE CAUSE OF PSEUDACHONDROPLASIA  
 CC (PSACH) AND MULTIPLE EPIPHYSEAL DYSPLASIA (MED OR EDW1), WHICH ARE  
 CC DOMINANTLY INHERITED CHONDRODYSPLASIAS CHARACTERIZED BY SHORT  
 CC STATURE AND EARLY-ONSET OSTEOARTHRITIS. MED IS BROADLY CATEGORIZED  
 CC INTO THE MORE SEVERE FAIRBANK AND THE Milder RIBBING TYPES. PSACH  
 CC IS MORE SEVERE AND IS RECOGNIZED IN EARLY CHILDHOOD.  
 CC -1 SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1 SIMILARITY: Contains 4 EGF-like domains.  
 CC -1 SIMILARITY: Contains 7 TSP type-3 domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).  
 CC or send an email to [licenses@isb-sib.ch](mailto:licenses@isb-sib.ch)).  
 CC -----  
 DR EMBL, I33137; AAA57253.1; -;  
 DR EMBL, S79499; AAB35269.1; -;  
 DR EMBL, S79500; AAB35270.1; -;  
 DR HSSP, P35444; 1VDF.  
 DR Genew, HGNC:2227; COMP.  
 DR MIM, 600310; -;  
 DR MIM, 132400; -;  
 DR MIM, 177170; -;  
 DR GO, GO:0005578; C:extracellular matrix; TAS.  
 DR GO, GO:0005509; F:calcium ion binding activity; TAS.  
 DR GO, GO:0005201; F:extracellular matrix structural constituent; TAS.  
 DR GO, GO:0007397; P:histogenesis and organogenesis; TAS.  
 DR GO, GO:0001501; P:skeletal development; TAS.  
 DR InterPro, IPR001681; EGF\_Ca.  
 DR InterPro, IPR006209; EGF\_Like.  
 DR InterPro, IPR003367; tsp\_3.  
 DR Pfam, PF00008; EGF\_2.  
 DR Pfam, PF02412; tsp\_3; 11.

DR	SMART, SM00179; EGF CA, 2.
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS01187; EGF CA, 2.
KM	Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain
KM	Signal; Disease mutation.
FT	SIGNAL 1 20 POTENTIAL.
FT	CHAIN 21 757 CARTILAGE OLIGOMERIC MATRIX PROTEIN.
FT	DOMAIN 22 86 N-TERMINAL.
FT	DOMAIN 87 126 EGF-LIKE 1.
FT	DOMAIN 127 179 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 180 222 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 225 267 EGF-LIKE 4.
FT	DOMAIN 297 332 TSP TYPE-3 1.
FT	DOMAIN 333 355 TSP TYPE-3 2.
FT	DOMAIN 356 391 TSP TYPE-3 3.
FT	DOMAIN 392 414 TSP TYPE-3 4.
FT	DOMAIN 415 452 TSP TYPE-3 5.
FT	DOMAIN 453 488 TSP TYPE-3 6.
FT	DOMAIN 489 524 TSP TYPE-3 7.
FT	DOMAIN 525 757 C-TERMINAL.
FT	DISULFID 69 69 INTERCHAIN (PROBABLE).
FT	DISULFID 72 72 INTERCHAIN (PROBABLE).
FT	DISULFID 91 102 BY SIMILARITY.
FT	DISULFID 96 111 BY SIMILARITY.
FT	DISULFID 114 125 BY SIMILARITY.
FT	DISULFID 131 142 BY SIMILARITY.
FT	DISULFID 136 151 BY SIMILARITY.
FT	DISULFID 154 178 BY SIMILARITY.
FT	DISULFID 184 197 BY SIMILARITY.
FT	DISULFID 191 206 BY SIMILARITY.
FT	DISULFID 209 221 BY SIMILARITY.
FT	DISULFID 229 243 BY SIMILARITY.
FT	DISULFID 237 253 BY SIMILARITY.
FT	DISULFID 255 266 BY SIMILARITY.
FT	DISULFID 282 287 BY SIMILARITY.
FT	DISULFID 292 312 BY SIMILARITY.
FT	DISULFID 328 348 BY SIMILARITY.
FT	DISULFID 351 371 BY SIMILARITY.
FT	DISULFID 387 407 BY SIMILARITY.
FT	DISULFID 410 430 BY SIMILARITY.
FT	DISULFID 448 468 BY SIMILARITY.
FT	DISULFID 484 504 BY SIMILARITY.
FT	DISULFID 520 741 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 742 742 CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE 367 369 D -> N (IN PSACH; MILD FORM).
FT	VARIANT 290 290 /FTID=VAR_007614.
FT	VARIANT 299 299 G -> R (IN PSACH; MILD FORM).
FT	VARIANT 328 328 /FTID=VAR_007615.
FT	VARIANT 342 342 C -> R (IN PSACH; MILD FORM).
FT	VARIANT 349 349 D -> Y (IN MED; FAIRBANK TYPE).
FT	VARIANT 361 361 /FTID=VAR_007617.
FT	VARIANT 361 361 D -> V (IN MED; FAIRBANK TYPE).
FT	VARIANT 361 361 D -> Y (IN MED).
FT	VARIANT 367 368 /FTID=VAR_007620.
FT	VARIANT 371 371 MISSING (IN MED).
FT	VARIANT 371 371 /FTID=VAR_007621.
FT	VARIANT 372 372 C -> S (IN MED; FAIRBANK TYPE).
FT	VARIANT 374 374 /FTID=VAR_007622.
FT	VARIANT 374 374 MISSING (IN PSACH).
FT	VARIANT 387 387 /FTID=VAR_007623.
FT	VARIANT 391 394 MISSING (IN PSACH; MILD FORM).
FT	VARIANT 391 394 /FTID=VAR_007624.
FT	VARIANT 391 394 C -> G (IN PSACH; MILD FORM).
FT	VARIANT 391 394 /FTID=VAR_007625.
FT	VARIANT 391 394 PMSD -> V (IN PSACH).
FT	VARIANT 391 394 /FTID=VAR_007626.

Query Match 44.9%; Score 262.5; DB 1; Length 757;  
 Best Local Similarity 69.5%; Pred. No. 2.7e-15;  
 Matches 57; Conservative 5; Mismatches 9; Indels 11; Gaps 2;

DB 10 L1L1AALGASGGGSPIG-SDLGPMRLRELOETNALQDVRLRQVREITFLKNT 67  
 10 L1L1AALGASGGGSPIG-SDLGPMRLRELOETNALQDVRLRQVREITFLKNT 65

QY 8 L1L1AALGASGGGSPIG-SDLGPMRLRELOETNALQDVRLRQVREITFLKNT 67  
 10 L1L1AALGASGGGSPIG-SDLGPMRLRELOETNALQDVRLRQVREITFLKNT 65

DB 66 VMECDACGMOQSVRTGSLVSR 87

RESULT 2  
 COMP RAT STANDARD; PRT; 755 AA.  
 AC P35444;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cartilage oligomeric matrix protein precursor (COMP).  
 GN COMP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RC TISSUE=Cartilage;  
 RA MEDLINE=93054522; PubMed=1429587;  
 RA Oldberg A., Antonsson P., Lindblom K., Heinegaard D.;  
 RT "COMP (cartilage oligomeric matrix protein) is structurally related  
 to the thrombospondins.";  
 RL J. Biol. Chem. 267:22346-22350(1992).  
 RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 27-72.  
 RX MEDLINE=97020114; PubMed=8864111;  
 RX Malashkevich V.N., Kammerer R.A., Efimov V.P., Schultness T.,  
 RA Engel J.;  
 RT "The crystal structure of a five-stranded coiled coil in COMP: a  
 RT prototype ion channel?";  
 RL Science 274:761-765(1996).  
 CC -1- SUBUNIT: PENTAMER; DISULFIDE-LINKED.  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1- SIMILARITY: Contains 4 EGF-like domains.  
 CC -1- SIMILARITY: Contains 7 TSP type-3 domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X72914; CAAS1419.1; -  
 CC PIR; A44315; A44315.  
 CC PDB; 1VDF; 08-OCT-97.  
 CC PDB; 1FBM; 09-AUG-00.  
 CC InterPro; IPR001881; EGF\_CA.  
 CC InterPro; IPR006209; EGF\_1like.  
 CC InterPro; IPR003367; tsp\_3.  
 CC Pfam; PF00008; EGF; 1.  
 CC Pfam; PF02412; tsp\_3; 9.  
 CC SMART; SM00179; EGF\_CA; 2.  
 CC PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 CC PROSITE; PS01186; EGF\_2; 1.  
 CC PROSITE; PS01187; EGF\_CA; 2.  
 CC Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;  
 CC Signal; 3D-structure.  
 CC SIGNAL 1 19 POTENTIAL.  
 CC CHAIN 20 755 CARTILAGE OLIGOMERIC MATRIX PROTEIN.  
 CC FT

FT DOMAIN. 21 84 N-TERMINAL.  
 FT DOMAIN 85 124 EGF-LIKE 1.  
 FT DOMAIN 125 177 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 178 220 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 223 265 EGF-LIKE 4.  
 FT DOMAIN 295 330 TSP TYPE-3 1.  
 FT DOMAIN 331 353 TSP TYPE-3 2.  
 FT DOMAIN 354 389 TSP TYPE-3 3.  
 FT DOMAIN 390 412 TSP TYPE-3 4.  
 FT DOMAIN 413 450 TSP TYPE-3 5.  
 FT DOMAIN 451 486 TSP TYPE-3 6.  
 FT DOMAIN 487 522 TSP TYPE-3 7.  
 FT DOMAIN 523 755 C-TERMINAL.  
 FT DISULFID 68 68 INTERCHAIN.  
 FT DISULFID 71 71 INTERCHAIN.  
 FT DISULFID 89 100 BY SIMILARITY.  
 FT DISULFID 94 109 BY SIMILARITY.  
 FT DISULFID 112 123 BY SIMILARITY.  
 FT DISULFID 129 140 BY SIMILARITY.  
 FT DISULFID 134 149 BY SIMILARITY.  
 FT DISULFID 152 176 BY SIMILARITY.  
 FT DISULFID 182 195 BY SIMILARITY.  
 FT DISULFID 189 204 BY SIMILARITY.  
 FT DISULFID 207 219 BY SIMILARITY.  
 FT DISULFID 227 241 BY SIMILARITY.  
 FT DISULFID 235 251 BY SIMILARITY.  
 FT DISULFID 253 264 BY SIMILARITY.  
 FT DISULFID 280 285 BY SIMILARITY.  
 FT DISULFID 290 310 BY SIMILARITY.  
 FT DISULFID 326 346 BY SIMILARITY.  
 FT DISULFID 349 369 BY SIMILARITY.  
 FT DISULFID 385 405 BY SIMILARITY.  
 FT DISULFID 408 428 BY SIMILARITY.  
 FT DISULFID 446 466 BY SIMILARITY.  
 FT DISULFID 482 502 BY SIMILARITY.  
 FT CARBOHYD 518 739 BY SIMILARITY.  
 FT CARBOHYD 740 740 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT HELIX 30 66 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT TURN 67 67  
 FT HELIX 69 71  
 FT HELIX 71 71  
 SQ SEQUENCE 755 AA; 82663 MW; AB48888FE093C598 CRC64;

Query Match 40.1%; Score 234; DB 1; Length 755;  
 Best Local Similarity 63.0%; Pred. No. 7.3e-13;  
 Matches 51; Conservative 8; Mismatches 18; Indels 4; Gaps 1;

DB 8 L1L1AALGASGGGSPIG-SDLGPMRLRELOETNALQDVRLRQVREITFLKNT 67  
 9 VL1AALRATGCGIPIAG-SDLGPMRLRELOETNALQDVRLRQVREITFLKNT 64

QY 68 VMECDACGMOQSVRTGSLVSR 88  
 65 VMECDACGMOQSVRTGSLVSR 85

DB 65 VMECDACGMOQSVRTGSLVSR 85

RESULT 3  
 ID TSP4 HUMAN STANDARD; PRT; 961 AA.  
 AC P35443;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Thrombospondin 4 precursor.  
 GN THBS4 OR TSP4.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RC TISSUE=Heart;  
 RX MEDLINE=93353522; PubMed=8350346;

RA	Lawler J., Duquette M., Urry L., McHenry K., Smith T.F.;
Rt	"The evolution of the thrombospondin gene family.";
RL	J. Mol. Evol. 36:509-516(1993).
CC	-1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN, LAMININ AND TYPE V COLLAGEN.
CC	-1- SUBUNIT: Homotrimer; disulfide-linked.
CC	-1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC	-1- SIMILARITY: Contains 4 EGF-like domains.
CC	-1- SIMILARITY: Contains 7 TSP type-3 domains.
CC	-1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-stb.ch/announce/">http://www.isb-stb.ch/announce/</a> or send an email to <a href="mailto:license@isb-stb.ch">license@isb-stb.ch</a> ).
CC	-----
DR	EMBL; Z19585; CAA79635..1; ..
DR	PIR; A55710; TSHUP4.
DR	HSSP; P35444; IYDF.
DR	Genew; HGNC:11786; THBS4.
DR	MIM; 600715; ..
DR	GO; GO:0005578; C:extracellular matrix; TAS.
DR	GO; GO:0005509; F:calcium ion binding activity; TAS.
DR	GO; GO:0008201; F:heparin binding activity; TAS.
DR	GO; GO:0006930; P:suberate-bound cell migration, cell extension; TAS.
DR	InterPro; IPR001881; EGF_Ca.
DR	InterPro; IPR006209; EGF_Like.
DR	InterPro; IPR003167; tcd_3.
DR	InterPro; IPR003129; TSPN.
DR	Pfam; PF00008; EGF; 2.
DR	Pfam; PF02412; tcd_3; 9.
DR	Pfam; PF02210; TSPN; 1.
DR	SMART; SM00179; EGF_CA; 2.
DR	SMART; SM00210; TSEN; 1.
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS01187; EGF_CA; 2.
KW	Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain; Signal.
FT	SIGNAL. 1 21 POTENTIAL.
FT	CHAIN 22 961 THROMBOSPONDIN 4.
FT	DOMAIN 22 285 N-TERMINAL.
FT	DOMAIN 286 325 EGF-LIKE 1.
FT	DOMAIN 326 378 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 379 419 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 420 462 EGF-LIKE 4.
FT	DOMAIN 492 527 TSP TYPE-3 1.
FT	DOMAIN 528 550 TSP TYPE-3 2.
FT	DOMAIN 551 586 TSP TYPE-3 3.
FT	DOMAIN 587 609 TSP TYPE-3 4.
FT	DOMAIN 610 647 TSP TYPE-3 5.
FT	DOMAIN 648 687 TSP TYPE-3 6.
FT	DOMAIN 688 723 TSP TYPE-3 7.
FT	C-TERMINAL.
FT	DOMAIN 724 961
FT	SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID 258 258 INTERCHAIN (PROBABLE).
FT	DISULFID 261 261 INTERCHAIN (PROBABLE).
FT	DISULFID 290 301 BY SIMILARITY.
FT	DISULFID 295 310 BY SIMILARITY.
FT	DISULFID 313 324 BY SIMILARITY.
FT	DISULFID 330 341 BY SIMILARITY.
FT	DISULFID 335 350 BY SIMILARITY.
FT	DISULFID 353 377 BY SIMILARITY.
FT	DISULFID 383 394 BY SIMILARITY.
FT	DISULFID 388 403 BY SIMILARITY.
FT	DISULFID 406 418 BY SIMILARITY.
FT	DISULFID 424 438 BY SIMILARITY.
FT	DISULFID 432 448 BY SIMILARITY.
FT	DISULFID 450 461 BY SIMILARITY.

```

FT DISULFID 477 482 BY SIMILARITY.
FT DISULFID 487 507 BY SIMILARITY.
FT DISULFID 523 543 BY SIMILARITY.
FT DISULFID 546 566 BY SIMILARITY.
FT DISULFID 582 602 BY SIMILARITY.
FT DISULFID 603 625 BY SIMILARITY.
FT DISULFID 643 663 BY SIMILARITY.
FT DISULFID 683 703 BY SIMILARITY.
FT DISULFID 719 940 BY SIMILARITY.
FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 941 941 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 961 AA; 105801 MW; 55A48AF481AE3DD CRC64;

Query Match 24.9%; Score 145.5; DB 1; Length 961;
Best Local Similarity 42.9%; Pred. No. 3,46-05;
Matches 30; Conservative 10; Mismatches 25; Indels 5; Gaps 2;

QY 31 DLGFQMLRELQETNALQDVYRWIRGQVREITFLKNTVMCEDCAGP---QPPKPPQPP 86
Db 218 DFNQSGFQGMQTQLNLQGLGEVADLIRQQVKETSPFRNTIAIECOAGCPKPFQSPSTVAVP 277
QY 87 QPPKPPQPP 96
Db 278 AP-PAPTRP 286

RESULT 4
TSP4 RAT
ID TSP4 RAT STANDARD; PRT; 980 AA.
AC P49744;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thrombospondin 4 precursor.
GN THBS4 OR TSP4 OR TSP-4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lewis; TISSUE=Skeletal muscle;
RX MEDLINE=96074771; PubMed=7490284;
RA Adler S., Catoni P.;
RT "Thrombospondin-4, an extracellular matrix protein expressed in the
RT developing and adult nervous system promotes neurite outgrowth.";
RL J. Cell Biol. 131:1083-1094(1995).
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN.
CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: Contains 4 EGF-like domains.
CC -1- SIMILARITY: Contains 7 TSP type-3 domains.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way that
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; X89963; CAA62002.1; -.
CC HSSP; P35444; 1VDF.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR003367; TSP_3.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF02412; TSP_3; 9.

```



DR Pfam; PF02210; TSPN; 1.  
 DR SMART; SM00179; EGF CA; 2.  
 DR SMART; SM00210; TSPN; 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF CA; 2.  
 KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;  
 KW Signal.

FT SIGNAL 1 39 POTENTIAL.  
 FT CHAIN 40 980 THROMBOSPONDIN 4.  
 FT DOMAIN 40 303 N-TERMINAL.  
 FT DOMAIN 304 343 EGF-LIKE 1.  
 FT DOMAIN 344 396 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 397 437 EGF-LIKE 3. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 438 481 EGF-LIKE 4.  
 FT DOMAIN 511 546 TSP TYPE-3 1.  
 FT DOMAIN 547 569 TSP TYPE-3 2.  
 FT DOMAIN 570 605 TSP TYPE-3 3.  
 FT DOMAIN 606 628 TSP TYPE-3 4.  
 FT DOMAIN 629 666 TSP TYPE-3 5.  
 FT DOMAIN 667 706 TSP TYPE-3 6.  
 FT DOMAIN 707 742 TSP TYPE-3 7.  
 FT DOMAIN 743 980 C-TERMINAL.  
 FT DISULFID 276 276 INTERCHAIN (PROBABLE).  
 FT DISULFID 279 279 INTERCHAIN (PROBABLE).  
 FT DISULFID 308 319 BY SIMILARITY.  
 FT DISULFID 313 328 BY SIMILARITY.  
 FT DISULFID 331 342 BY SIMILARITY.  
 FT DISULFID 348 359 BY SIMILARITY.  
 FT DISULFID 353 368 BY SIMILARITY.  
 FT DISULFID 371 395 BY SIMILARITY.  
 FT DISULFID 401 412 BY SIMILARITY.  
 FT DISULFID 406 421 BY SIMILARITY.  
 FT DISULFID 424 436 BY SIMILARITY.  
 FT DISULFID 442 456 BY SIMILARITY.  
 FT DISULFID 450 466 BY SIMILARITY.  
 FT DISULFID 468 480 BY SIMILARITY.  
 FT DISULFID 496 501 BY SIMILARITY.  
 FT DISULFID 506 526 BY SIMILARITY.  
 FT DISULFID 542 562 BY SIMILARITY.  
 FT DISULFID 565 585 BY SIMILARITY.  
 FT DISULFID 601 621 BY SIMILARITY.  
 FT DISULFID 624 644 BY SIMILARITY.  
 FT DISULFID 662 682 BY SIMILARITY.  
 FT DISULFID 702 722 BY SIMILARITY.  
 FT DISULFID 738 959 BY SIMILARITY.  
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 980 AA; 108213 MW; 056D41EB6E206FCF CRC64;

Query Match 24.7%; Score 144.5; DB 1; Length 980;  
 Best Local Similarity 42.0%; Pred. No. 4.2e-05;  
 Matches 29; Conservative 10; Mismatches 27; Indels 3; Gaps 1;

QY 31 DLGPMRELEQETAAALQDVDMRLQVREITPKNTWECADGPPQPPQPPQ---PQ 87  
 DB 236 DFNQPFGLGQMQLNLGVEVDLRLQVKEKSPFLNTIAECQACGSPFSQSPNTLVPI 295

QY 88 PPKPPKPP 96  
 DB 296 APPAPPPR 304

RESULT 5  
 TSP4\_XENLA STANDARD; PRT; 955 AA.  
 AC Q06441;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 28, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Thrombospondin 4 precursor.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93163109; PubMed=8432726;  
 RA Lawler J., Duguet M., Whitaker C.A., Adams J.C., McHenry K.,  
 RA Desimone D.W.;  
 RT "Identification and characterization of thrombospondin-4, a new  
 member of the thrombospondin gene family.";  
 RL J. Cell Biol. 120:1059-1067(1993).  
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
 LAMININ AND TYPE V COLLAGEN. MAY PARTICIPATE IN THE GENESIS AND  
 FUNCTION OF CARDIAC AND SKELETAL MUSCLE.  
 CC -1- SUBUNIT: Homotrimer; disulfide-linked.  
 CC -1- DEVELOPMENTAL STAGE: INITIAL EXPRESSION DURING NEURULATION.  
 CC INCREASE DURING TALIBD STAGES BUT DECREASE BY THE FEEDING TADPOLE  
 STAGE.  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1- SIMILARITY: Contains 4 EGF-like domains.  
 CC -1- SIMILARITY: Contains 7 TSP type-3 domains.  
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z19091; CA79518.1; -.  
 DR PIR; A45441; A45441.  
 DR HSSP; P35444; IVDF.  
 DR InterPro; IPR001881; EGF CA.  
 DR InterPro; IPR006209; EGF like.  
 DR InterPro; IPR003367; TSP 3.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF00008; EGF; 2.  
 DR Pfam; PF02412; TSP 3; 9.  
 DR Pfam; PF02210; TSPN; 1.  
 DR Pfam; PF02210; TSPN; 1.  
 DR SMART; SM00179; EGF CA; 2.  
 DR SMART; SM00210; TSPN; 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF CA; 2.  
 KW Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;  
 KW Signal.

FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 955 THROMBOSPONDIN 4.  
 FT DOMAIN 25 280 N-TERMINAL.  
 FT DOMAIN 281 320 EGF-LIKE 1.  
 FT DOMAIN 321 373 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 374 417 EGF-LIKE 3. CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 418 459 EGF-LIKE 4.  
 FT DOMAIN 489 524 TSP TYPE-3 1.  
 FT DOMAIN 525 547 TSP TYPE-3 2.  
 FT DOMAIN 548 583 TSP TYPE-3 3.  
 FT DOMAIN 584 606 TSP TYPE-3 4.  
 FT DOMAIN 607 644 TSP TYPE-3 5.  
 FT DOMAIN 645 684 TSP TYPE-3 6.  
 FT DOMAIN 685 720 TSP TYPE-3 7.  
 FT DOMAIN 714 955 C-TERMINAL.  
 FT DISULFID 256 256 INTERCHAIN (PROBABLE).  
 FT DISULFID 259 259 INTERCHAIN (PROBABLE).  
 FT DISULFID 285 296 BY SIMILARITY.  
 FT DISULFID 290 305 BY SIMILARITY.  
 FT DISULFID 308 319 BY SIMILARITY.  
 FT DISULFID 325 336 BY SIMILARITY.  
 FT DISULFID 330 345 BY SIMILARITY.  
 FT DISULFID 348 372 BY SIMILARITY.  
 FT DISULFID 378 392 BY SIMILARITY.



```

Oy      76  P O P O K P O P O P K P O P K P E P E T G S S E K D 1 0 7
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      224  P O P O P O P O P O P O P O P O P O P O G G A N N N K N N 255

RESULT 8
TEGU_HSV11
ID      TEGU_HSV11      STANDARD;      PRT;      3164 AA.
AC      P10270;
DT      01-MAR-1989 (Rel. 10, Created)
DT      01-MAR-1989 (Rel. 10, Last sequence update)
DT      01-APR-1993 (Rel. 25, Last annotation update)
DE      Large tegument protein (Virion protein UL36).
GN      UL36.
OS      Herpes simplex virus (type 1 / strain 17).
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Alphaherpesvirinae; Simplexvirus.
OX      NCBI_TaxID=10299;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      MEDLINE=88274327; PubMed=2839594;
RA      McGeoch D.J., Dairymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA      McNab D., Perry L.J., Scott J.E., Taylor P.,
RA      "The complete DNA sequence of the long unique region in the genome of
RT      herpes simplex virus type 1."
RL      J. Gen. Virol. 69:1531-1574(1988).
CC      - FUNCTION: TEGUMENT PROTEIN.
CC      - SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC      EBV-1 24, EBV BLF1, HVS-1 64, VAV 22, AND HCMV UL48.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X14112; CAA32311.1; -.
DR      PIR; I30085; WMBEH6.
DR      InterPro; IPR006928; Herpes teg N.
DR      InterPro; IPR005210; Herpes_UL36.
DR      Pfam; PF04843; Herpes_teg_N; 1.
DR      Pfam; PF03586; Herpes_UL36; 1.
KW      Repeat.
KW      DOMAIN
FT      DOMAIN          2911      2980          35 X 2 AA TANDEM REPEATS OF P-Q.
SQ      SEQUENCE      3164 AA; 335857 MW; CCS031PF4PFE3F4 CRG4;

Query Match      22.6%; Score 132; DB 1; Length 3164;
Best Local Similarity 66.7%; Pred. NO. 0.0016;
Matches 22; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Oy      67  T V M E C D A C G P O P K P O P O P K P O P K P E P E 99
          |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      2904  T V S R L S A P O P O P O P O P O P O P O P O P O P O Q 2936

RESULT 9
NOL3_MOUSE
ID      NOL3_MOUSE      STANDARD;      PRT;      220 AA.
AC      Q9D1X0;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Nucleolar protein 3.
GN      NOL3.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      STRAIN=C57BL/6J; TISSUE=Adipose tissue;

```

RX	MEDLINE:21085660; PubMed=11217851.
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA	Arikawa T., Hara A., Fukunishi Y., Komio H., Adachi J., Fukuda S.,
RA	Atzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiba H.,
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA	Schirra L.W., Staabli F., Suzuki K., Tomita M., Wagner U., Washio T.,
RA	Sasai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA	Gustincich S., Hill D., Hofmann M., Hunne D.A., Kamiya M., Lee N.H.,
RA	L Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA	Suzuki H., Toyooka K., Wang K.H., Welte C., Whitaker C., Wilming L.,
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsauki S.,
RA	Hayashizaki Y.,
RT	"Functional annotation of a full-length mouse cDNA collection.";
RL	Nature 409:685-690(2001).
CC	-1 FUNCTION: May be involved in RNA splicing (By similarity).
CC	-1 SUBUNIT: Interacts with SRP30c, NPM1, CASP2, CASP8 and CED-3 (By
CC	similarity).
CC	-1 SUBCELLULAR LOCATION: Nuclear (By similarity).
CC	-1 SIMILARITY: Contains 1 CARD domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL: AK021023; BAB32281.1; --
DR	MED; MG1:1925938; NC013
DR	InterPro; IPR001315; CARD.
DR	SMART; SM00114; CARD; 1.
DR	PROSITE; PS50209; CARD; 1.
KW	Nuclear protein; mRNA splicing.
FT	DOMAIN 4 95 CARD.
FT	132      218     GLU/PRO-RICH.
SO	SEQUENCE 220 AA; 24567 MW; A4DCD57C1EB320A2 CRC64;
Query Match	20.0%; Score 117; DB 1; Length 220;
Best Local Similarity	30.7%; Pred. No. 0.0021;
Matches 31; Conservative 15; Mismatches 25; Indels 30; Gaps 5;	
Oy	26 GDCCSDI-----GPOMLRELQ-----ETNAALQDVRLRQOVREITFLKN 66
Dy	121 GTTCPELP.PRASGEVGGPGSSSLAQRTPEPEPELEAEATGDEPD-LQDEM-----N 172
Oy	67 TWNECDACGPDPKPPOPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPP 107
Dy	173 PEQEPE---PEPD 210
RESULT 10	
ID	TOMB_HELPJ STANDARD; PRT; 280 AA.
AC	Q9ZJP4;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Tomb protein.
GN	TOMB OR JHP1260.
OS	Helicobacter pylori J99 (Campylobacter pylori J99).
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC	Helicobacteraceae; Helicobacter.
OX	NCBI_TaxID=85963;
RV	[1]
RP	SEQUENCE FROM N.A.

```

RA XM R.A., Ling L.-S.L., Mofir D.T., King B.L., Brown E.D., Dolg P.C.,
RA Alm R.A., 99120557, PubMed=9923682.
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -I- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENER-
CC GY-REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC MEMBRANE PROTEINS (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001549; AACD06833.1; -.
DR F1R, F71829, F71829.
DR InterPro; IPR003538; TonB.
DR InterPro; IPR006260; TonB_C.
DR Pfam; PF03544; TonB_1.
DR PRINTS; PR01374; TONBPROTEIN.
DR TRIGPAMS; TIGR01352; tonB Ctem; 1.
KW Transport, Protein transport, Inner membrane, Periplasmic;
KW Transmembrane, Signal-anchor, Repeat, Complete proteome.
FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 15 35 SIGNAL-ANCHOR (POTENTIAL).
FT PT DOMAIN 36 280 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 280 AA; 31045 MW; 640717A8BF81D76D CRC64;

Query Match 19.7%; Score 115; DB 1; Length 280;
Best Local Similarity 54.5%; Pred. No. 0.0039;
Matches 18; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

Qy 76 PQPQPQPQPQPQPQPQPQPKEPETGSSSEKDE 108
Db 99 PKPKPKPKPKPKPKPKPKPKPKPEKPYEKPKE 131

RESULT 11
TONB_HELPY
ID TONB_HELPY STANDARD; PRT; 285 AA.
AC O25899;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TonB protein.
GN TONB OR HP1341.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
CX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695; ATCC 700392;
RX MEDLINE=97394467; PubMed=9925185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson D., Khalak H.G., Glodek A.,
```

```

RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickley E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Welden J.W., Fujii C., Bowman C., Watney L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997)
CC CC
CC -|- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC MEMBRANE PROTEINS (BY SIMILARITY).
CC CC
CC -|- SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC CC
CC -|- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000635; AAD08383.1; -.
DR PIR; E64687; E64687.
DR TIGR; HP1341;
DR InterPro; IPR003538; TonB.
DR InterPro; IPR006260; TonB_C.
DR Pfam; PF03544; TonB_1.
DR PRINTS; PR01374; TONBPOTREIN.
DR TRIGRFAMS; TIGR01352; tonB.Cterm; 1.
KM Transport; Protein transport; Inner membrane; Periplasmic;
KT Transmembrane; Signal-anchor; Repeat; Complete proteome.
FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 15 35 SIGNAL-ANCHOR (POTENTIAL).
FT FT DOMAIN 36 285 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 285 AA; 31585 MW; 3897A8BB7B70BDP8 CRC64;

Query Match 19.7%; Score 115; DB 1; Length 285;
Best local Similarity 66.7%; Pred. No. 0.004;
Matches 16; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 76 POPPKPPKPPKPPKPPKPPKE 99
|::|::|::|::|::|::|::|::|:
Db 106 PKPKPKPKPKPKPKPKPKPKPK 129

RESULT 12
TRA9_MAIZE STANDARD; PRT; 839 AA.
AC P03010;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Putative AC9 transposase.
DS Zee maye (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84205699; Pubmed=6327080;
RA Pohlman R.F., Fedoroff N.V., Messing J.;
RT "The nucleotide sequence of the maize controlling element Activator.";
RL Cell 37:635-643(1984).
CC -|- MISCELLANEOUS: THIS PROTEIN IS CODED BY THE TRANSPOSABLE MAIZE
CC CONTROLLING ELEMENT "ACTIVATOR" (AC), WHICH IS ABLE TO ACTIVATE

```

	CC	CHROMOSOME BREAKAGE AT A SPECIFIC LOCATION; IT MAY BE THE
	CC	STRUCTURAL GENE FOR A TRANS-ACTING FUNCTION REQUIRED FOR
	CC	TRANSPOSITION.
	CC	- - - - -
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
	CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
	CC	the European Bioinformatics Institute. There are no restrictions on its
	CC	use by non-profit institutions as long as its content is in no way
	CC	modified and this statement is not removed. Usage by and for commercial
	CC	entities requires a license agreement (See http://www.isib.ch/announce/
	CC	or send an email to license@isib-sib.ch).
	CC	- - - - -
	DR	EMBL; X01964.; NOT_ANNOTATED_CDS.
	DR	PIR; K03540; TOZMCA.
	DR	MazzeDB; 65747.-.
	DR	InterPro; IPR003656; BED_finger.
	DR	Pfam; PF02892; zf-BED; 1.
	DR	SMART; SM00614; Znf_BED; 1.
KW	Transposable element; Transposition; DNA-binding; DNA recombination.	
SQ	SEQUENCE    939 AA; 97068 MW; 803606AA325EA042 CRC64;	
Oy	56 QOYREITFLNKTVCEACGCP--QPQPKPQQPPQPKPCKPE 99                   :                  Db    19 QQLRREHFLOSVSSSNANGNTATPSQDDMAIVHEPOPOPFPFEPDOPPFEEBEE 75	
RESULT 13		
ID	PARX TRYBB	STANDARD; PRT; 115 AA.
AC	pI40z3;	
DT	01-APR-1990 (Rel. 14, Created)	
DT	01-APR-1990 (Rel. 14, Last sequence update)	
DE	15-JUL-1999 (Rel. 38, last annotation update)	
DJ	Procyclic form specific polypeptide precursor (Procyclin) (PARP).	
GN	PROA.	
OS	Trypanosoma brucei brucei.	
OC	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.	
OX	NCB1_taxid=5702;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=227(ITTAT1);	
RX	MEDLINE=90067841; PubMed=2573878;	
RA	Xeenig E., Delius H., Carrington M., Williams R.O., Roditi I.;	
RT	"Duplication and transcription of procyclin genes in Trypanosoma brucae".	
RL	Nucleic Acids Res. 17:8727-8739(1989).	
CC	--FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS. --SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor. --DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING DIFFERENTIATION IN THE INSECT VECTOR.	
CC	- - - - -	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isib-sib.ch/announce/ or send an email to license@isib-sib.ch).	
CC	- - - - -	
CR	EMBL; X16015; CAA34147.L; --	
DR	PIR; SIA896; PAUTPC.	
KW	Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.	
FT	SIGNAL          1            27	
FT	CHAIN           28          93	PROCYCLOC FORM SPECIFIC POLYPEPTIDE.
FT	PROPEP          94          115	
FT	DOMAIN          59          92	17 x 2 AA TANDEM REPEATS OF [DE]-P-
FT	CARBOHYD        56          56	N-LINKED (GLCNAC.....)(POTENTIAL).
SQ	SEQUENCE    115 AA; 11714 MW; 6039FD731CF0509E CRC64;	

```

Query March 19.04; Score 111; DB 1; Length 115;
Best Local Similarity 58.3%; Prod. No. 0.0035;
Matches .14; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

Cy 76 POPQKPEPOPOPQKPEPEPE 99
|:::|:::|:::|:::|
Db 62 PEPEPEPEPEPEPEPEPEPEPE 85

RESULT 14
PARB_TRYBB STANDARD; PRT; 129 AA.
AC P09731;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Procyelic form specific polypeptide A-beta precursor (Procyclin) (PAP
A-beta).
GN PARPA-BETA.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
NC NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=427;
RX MEDLINE=87115776; PubMed=3808022.
RA Roditi I., Garrington M., Turner M.;
RT "Expression of a polypeptide containing a dipeptide repeat is
confined to the insect stage of Trypanosoma brucei.";
RL Nature 325:272-274(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258895; PubMed=2342468;
RA Clayton C.E., Peierl J.P., Itzhaki J.E., Belliofatto V., Sherman D.R.,
RA Wisdom G.S., Vijayasarathy S., Mowatt M.R.;
RT "Transcription of the procyelic acidic repetitive protein genes of
Trypanosoma brucei.";
RL Mol. Cell. Biol. 10:3036-3047(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=427;
RA Vijayasarathy S., Ernest I., Itzhaki J., Sherman D., Mowatt M.R.,
RA Michaels P.A.M., Clayton C.E.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE OF 28-64, AND POST-TRANSLATIONAL MODIFICATIONS.
RC STRAIN=427;
RX MEDLINE=89359323; PubMed=2475493;
RA Clayton C.E., Mowatt M.R.;
RT "The procyelic acidic repetitive proteins of Trypanosoma brucei.
Purification and post-translational modification.";
RL J. Biol. Chem. 264:15088-15093(1989).
RN -1- FUNCTION. MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING
DIFFERENTIATION IN THE INSECT VECTOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X04814; CAAG28503.1; -
DR EMBL, M33129; AAA30225.1; -
DR EMBL, X52584; CAA36815.1; -
DR PIR, A26036; A26036.
KW Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.
FT SIGNAL 1
FT CHAIN 28 107 PROCYCLIC FORM SPECIFIC POLYPEPTIDE A-
BT BETA.
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:04:38 ; Search time 117.681 Seconds  
(without alignments)  
239.016 Million cell updates/sec

Title: US-09-696-872-23  
Perfect score: 584  
Sequence: 1 MRVWILGLALAAVCSAAK.....PKPQKPEPEGTSGSEKDEL 109

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	256.5	43.9	130	4	Q8N2R4
2	247.5	42.4	755	6	Q9BG80
3	245.5	42.0	817	4	014592
4	236	40.4	755	11	Q9R0G6
5	236	40.4	755	11	Q8V154
6	195.5	33.5	724	4	Q8N4T2
7	153	26.2	1319	4	Q94885
8	146.5	25.1	863	11	Q9QY53
9	146.5	25.1	963	11	Q921T2
10	133	22.8	272	5	Q25648
11	133	22.8	332	5	Q25649
12	132	22.6	3164	12	Q69088
13	127	21.7	2387	16	Q8FFP8
14	126	21.6	416	5	Q03752
15	124	21.2	910	11	Q9JLE8
16	124	21.2	912	11	Q9JLE7

17	124	21.2	1494	11	Q88902	Q88902 rattus norv
18	123	21.1	823	11	Q9E0F1	Q9E0F1 rattus norv
19	123	21.1	989	11	Q9JLE9	Q9JLE9 rattus norv
20	122	20.9	259	4	Q8W24	Q8W24 homo sapien
21	122	20.9	378	2	Q9LB09	Q9LB09 vibrio para
22	122	20.9	400	4	Q15184	Q15184 homo sapien
23	121.5	20.8	170	10	Q9M5X3	Q9M5X3 oryza sativ
24	121	20.7	378	10	Q9ZNY1	Q9ZNY1 zea mays (m
25	121	20.7	675	11	Q9EPT3	Q9EPT3 mus musculu
26	121	20.7	704	11	Q9JLQ4	Q9JLQ4 mus musculu
27	120.5	20.6	565	5	Q15754	Q15754 dictyosteli
28	120	20.5	273	10	Q9ARY7	Q9ARY7 oryza sativ
29	120	20.5	277	5	Q27045	Q27045 theileria p
30	120	20.5	374	2	Q33899	Q33899 streptococc
31	120	20.5	376	2	Q54071	Q54071 streptococc
32	120	20.5	377	2	Q9S6L9	Q9S6L9 streptococc
33	119	20.4	104	12	Q8V9Y2	Q8V9Y2 chorisoneu
34	119	20.4	164	10	Q8LUL1	Q8LUL1 saccharum h
35	119	20.4	230	10	Q8S4T0	Q8S4T0 saccharum h
36	119	20.4	667	11	Q35745	Q35745 mus musculu
37	118.5	20.3	374	5	Q8T4N5	Q8T4N5 glosina pa
38	118	20.2	165	5	Q22168	Q22168 caenorhabdi
39	118	20.2	422	16	Q8X3E0	Q8X3E0 escherichia
40	118	20.2	2344	5	Q9N3Y8	Q9N3Y8 caenorhabdi
41	117	20.0	220	11	Q8R2S3	Q8R2S3 mus musculu
42	117	20.0	220	11	Q8C550	Q8C550 mus musculu
43	117	20.0	309	12	Q89402	Q89402 paramecium
44	117	20.0	358	10	Q8W265	Q8W265 oryza sativ
45	117	20.0	480	5	Q27033	Q27033 theileria p

## ALIGNMENTS

RESULT 1

Q8N2R4 ID Q8N2R4 PRELIMINARY; PRT; 130 AA.

AC Q8N2R4; 01-OCT-2002 (TREMUR1, 22, Created)

DT 01-OCT-2002 (TREMUR1, 22, Last sequence update)

DT 01-OCT-2002 (TREMUR1, 22, Last annotation update)

DE Hypothetical protein FLJ90027.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RA Ieogal T., Oca T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T., Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagai K., Maehuo Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N., Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;

RA "NEDO human cDNA sequencing project."

RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

RL EMBL AK074508; BAC11031.1; .

DR EMBL AK074508; BAC11031.1; .

KW Hypothetical protein.

SO SEQUENCE 130 AA; 13832 MW; BFE7E5EBE7BACDD CRC64;

Query Match 43.9%; Score 256.5; DB 4; Length 130;  
Best Local Similarity 56.0%; Pred. No. 1.6e-18;  
Matches 61; Conservative 8; Mismatches 23; Indels 17; Gaps 3;

QY 8 LLAALAAVCSAAKSSIGGDCSDLGQMLRELQETNALQDVRLQGVREITFLKNT 67  
|||:::|  
Db 10 LTLTALAGAGGGGSPG---SDLGQMLRELQETNALQDVRLQGVREITFLKNT 65  
|||:::|

QY 68 VMECDACGPO-----PQPKPQPPQPP-----QPKPKPEPEGTSGS 103  
|||:::|  
Db 66 VMECDACGMOOSVRTGLPSVPRPLHCAFGVPLRGLPAGVGRPHPPGGA 114  
|||:::|







```

RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Weremowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Lawler J.;
RT "The thombospondin-4 gene.";
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF152393; AAD32714.1; -.
DR EMBL; AF152392; AAD32714.1; JOINED.
DR HSSP; P35444; 1VDF.
DR MGD; MGI:110179; Tnbs4.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR003167; tsp_3.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF02410; TSPN; 1.
DR Pfam; PF02412; tsp_3; 9.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00219; EGF_CA; 2.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW EGF-like domain.
FT NON TER
SQ SEQUENCE 863 AA; 95339 MW; 68E3EE6846728E4D CRC64;

Query Match 25.1%; Score 146.5; DB 11; Length 863;
Best Local Similarity 37.0%; Pred. No. 1.8e-06;
Matches 30; Conservative 12; Mismatches 36; Indels 3; Gaps 1;

Qy 19 KKGSSLGDCDDLPQMLRELQETNALQDVRDLRQGVREITFLKNTWECDCGCPQ 78
Db 108 QQSEPLAATSTGDFNRQFLGQMTQLNLGVEVDLRLQGVKETSFLNTIAECQACGPLS 167

Qy 79 QPKPQPQ---PQPKPQPQPKP 96
Db 168 FQSPPTNTLVPIAPPAPPTRP 188

RESULT 9
Q921T2 PRELIMINARY; PRT; 963 AA.
AC Q921T2;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Thombospondin-4.
GN THBS4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
OX NCBI_TaxID=10090;
RN NCI_TaxID=10090;
RP SEQUENCE FROM N.A.
RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Weremowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
RA Copeland N.G., Lawler J.;
RT "The thombospondin-4 gene.";
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF102887; AAC73003.1; -.
DR HSSP; P35444; 1VDF.
DR MGD; MGI:110179; Tnbs4.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR003167; tsp_3.
DR Pfam; PF00008; EGF; 1.
DR Pfam; PF02410; TSPN; 1.
DR Pfam; PF02412; tsp_3; 9.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00018; EF_HAND; 1.

```

```

DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW EGF-like domain.
SQ SEQUENCE 963 AA; 106366 MW; B8BA83B84F489FB1 CRC64;

Query Match 25.1%; Score 146.5; DB 11; Length 963;
Best Local Similarity 37.0%; Pred. No. 2e-06;
Matches 30; Conservative 12; Mismatches 36; Indels 3; Gaps 1;

Qy 19 KKGSSLGDCDDLPQMLRELQETNALQDVRDLRQGVREITFLKNTWECDCGCPQ 78
Db 208 QQSEPLAATSTGDFNRQFLGQMTQLNLGVEVDLRLQGVKETSFLNTIAECQACGPLS 267

Qy 79 QPKPQPQ---PQPKPQPQPKP 96
Db 268 FQSPPTNTLVPIAPPAPPTRP 288

RESULT 10
Q25648 PRELIMINARY; PRT; 272 AA.
AC Q25648;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Circumsporozoit (CS) protein (fragment).
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN NCI_TaxID=5821;
RP SEQUENCE FROM N.A.
RX MEDLINE=87218962; PubMed=3556207;
RA Weber J.L., Egan J.E., Lyon J.A., Wirtz R.A., Charoenvit Y.,
RA Maloy W.L., Hockmeyer W.T.;
RT "Plasmodium berghei: Cloning of the circumsporozoit protein gene.";
RL Exp. Parasitol. 63:295-300(1987).
DR EMBL; M25445; AAA29531.1; -.
DR InterPro; IPR003067; Circmeprozite.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp_1; 1.
DR PRINTS; PR01303; CIRCSPRZOTE.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
FT NON TER
SQ SEQUENCE 272 AA; 29408 MW; 4FF07FA62B32A051 CRC64;

Query Match 22.8%; Score 133; DB 5; Length 272;
Best Local Similarity 62.5%; Pred. No. 1.2e-05;
Matches 20; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

Qy 76 PQPKPQPQPKPQPKPPEPGTGSSEKD 107
Db 149 PQPKPQPQPKPQPKPQPKPGGNNNNKVN 180

RESULT 11
Q25649 PRELIMINARY; PRT; 332 AA.
AC Q25649;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Circumsporozoit protein.
GN CSP.
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN NCI_TaxID=5821;
RP SEQUENCE FROM N.A.
RX STRAIN=NK65;
RX MEDLINE=90158693; PubMed=2406593;

```

RA Lanar D.E.;  
 RT "Sequence of the circumsporozoite gene of Plasmodium berghel ANKA  
 RL clone and NK65 strain.";  
 RL Mol. Biochem. Parasitol. 39:151-154(1990).  
 DR EMBL; M28887; AAA29541.1; -.  
 DR InterPro; IPR003067; CircmSPRzote.  
 DR InterPro; IPR002965; P\_rich\_extenm.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; tsp\_1; 1.  
 DR PRINTS; PR01303; CRCMSPRZOTE.  
 DR PRINTS; PR01217; PRICHEXTENM.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 SQ SEQUENCE 332 AA; 36196 MW; 07C2D2F265EAAAB CRC64;

Query Match 22.8%; Score 133; DB 5; Length 332;  
 Best Local Similarity 62.5%; Pred. No. 1.5e-05;  
 Matches 20; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

Oy 76 PQPQPKPQPPQPKPQPKPEPGTGSSEKD 107  
 Db 209 PQPQPKPQPPQPKPQPKPEPGTGSSEKD 240

## RESULT 12

069088 PRELIMINARY; PRT; 3164 AA.  
 AC Q69088;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Viton protein.  
 OS Human herpesvirus 1.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10298;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17;  
 RX MEDLINE=86274327; PubMed=2839594;  
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,  
 RA McNaab D., Perry L.J., Scott J.E., Taylor P.;  
 RT "The complete DNA sequence of the long unique region in the genome of  
 RT herpes simplex virus type 1."  
 RL J. Gen. Virol. 69:1531-1574(1988).  
 DR EMBL; D10879; BA01682.1; -.  
 DR InterPro; IPR006928; Herpes\_leg\_N.  
 DR InterPro; IPR005210; Herpes\_leg\_N; 1.  
 DR Pfam; PF04843; Herpes\_leg\_N; 1.  
 DR Pfam; PF03586; Herpes\_UL36; 1.  
 SQ SEQUENCE 3164 AA; 335817 MW; 2E3335F9525F8C71 CRC64;

Query Match 22.6%; Score 132; DB 12; Length 3164;  
 Best Local Similarity 66.7%; Pred. No. 0.00021;  
 Matches 22; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Oy 67 TMECDACGPPQPKPQPKPQPKPEPE 99  
 Db 2304 TVSRSLAPQPPQPKPQPKPQPKPEPE 2936

## RESULT 13

08FF8 PRELIMINARY; PRT; 2387 AA.  
 ID Q8FF8;  
 AC Q8FF8;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Yaph homolog.  
 GN C2895.  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.

OX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:HI / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 DR EMBL; AB016764; AAN81345.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 2387 AA; 242283 MW; 06298DC5293DD03 CRC64;

Query Match 21.7%; Score 127; DB 16; Length 2387;  
 Best Local Similarity 76.9%; Pred. No. 0.0005;  
 Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy 76 PQPQPKPQPPQPKPQPKPEPGT 101  
 Db 2291 PQPQPKPQPPQPKPQPKPEPGT 2316

## RESULT 14

003752 PRELIMINARY; PRT; 416 AA.  
 AC Q03752;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Circumsporozoite protein.  
 GN CS.  
 OS Plasmodium yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5861;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9114845; PubMed=2290446;  
 RA Colomer-Gould V., Enea V.;  
 RT "Plasmodium yoelii nigeriensis circumsporozoite gene structure and its  
 RT implications for the evolution of the repeat regions."  
 RL Mol. Biochem. Parasitol. 43:51-58(1990).  
 DR EMBL; M58295; AAA29528.1; -.  
 DR InterPro; IPR003067; CircmSPRzote.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; tsp\_1; 1.  
 DR PRINTS; PR01303; CRCMSPRZOTE.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 SQ SEQUENCE 416 AA; 45135 MW; BD8C6AD168EBB7F0 CRC64;

Query Match 21.6%; Score 126; DB 5; Length 416;  
 Best Local Similarity 79.2%; Pred. No. 9.8e-05;  
 Matches 19; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Oy 76 PQPQPKPQPPQPKPQPKPEPE 99  
 Db 280 PQPQPKPQPPQPKPQPKPEPE 303

## RESULT 15

09ULE8 PRELIMINARY; PRT; 910 AA.  
 ID Q9ULE8;  
 AC Q9ULE8;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE GABA-A receptor epsilon-like subunit.  
 GN GABRE OR EPSILON.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 15:53:48 ; Search time 4.28319 Seconds  
(without alignments)  
148.232 Million cell updates/sec

Title: US-09-696-872-37

Perfect score: 20  
Sequence: 1 KDELL 4

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	20	100.0	4	14	AA1971
2	20	100.0	4	15	AA1971
3	20	100.0	4	15	AA1971
4	20	100.0	4	15	AA1971
5	20	100.0	4	15	AA1971
6	20	100.0	4	17	AA1971
7	20	100.0	4	17	AA1971
8	20	100.0	4	18	AA1971
9	20	100.0	4	18	AA1971

10	20	100.0	4	19	AA1971
11	20	100.0	4	19	AA1971
12	20	100.0	4	19	AA1971
13	20	100.0	4	19	AA1971
14	20	100.0	4	19	AA1971
15	20	100.0	4	19	AA1971
16	20	100.0	4	19	AA1971
17	20	100.0	4	19	AA1971
18	20	100.0	4	20	AA1971
19	20	100.0	4	20	AA1971
20	20	100.0	4	20	AA1971
21	20	100.0	4	20	AA1971
22	20	100.0	4	20	AA1971
23	20	100.0	4	20	AA1971
24	20	100.0	4	20	AA1971
25	20	100.0	4	20	AA1971
26	20	100.0	4	20	AA1971
27	20	100.0	4	20	AA1971
28	20	100.0	4	21	AA1971
29	20	100.0	4	21	AA1971
30	20	100.0	4	21	AA1971
31	20	100.0	4	21	AA1971
32	20	100.0	4	21	AA1971
33	20	100.0	4	21	AA1971
34	20	100.0	4	21	AA1971
35	20	100.0	4	21	AA1971
36	20	100.0	4	21	AA1971
37	20	100.0	4	21	AA1971
38	20	100.0	4	22	AA1971
39	20	100.0	4	22	AA1971
40	20	100.0	4	22	AA1971
41	20	100.0	4	22	AA1971
42	20	100.0	4	22	AA1971
43	20	100.0	4	22	AA1971
44	20	100.0	4	22	AA1971
45	20	100.0	4	22	AA1971

#### ALIGNMENTS

RESULT 1  
AA1971  
ID: AA1971 standard; peptide; 4 AA.  
AC: AA1971;  
XX: 25-MAR-2003 (updated)  
DT: 10-MAR-2003 (updated)  
DT: 22-OCT-1993 (first entry)  
XX: Mammalian endoplasmic reticulum retention signal.  
DE: yeast expression cassette; yeast alpha-factor leader; KEX2p;  
KW: heterologous protein production; mating pheromone; YAP3;  
KW: diabolic processing endoprotease.  
XX: Mammalia.  
OS: EP548012-A1.  
XX: 23-JUN-1993.  
PD: 08-DEC-1992; 92EP-0810964.  
XX: 16-DEC-1991; 91EP-0810984.  
PR: (CIBA ) CIBA GEIGY AG.  
PA: (NOVS ) NOVARTIS AG.  
XX: Chaudhuri B, Riezman H, Seebach P, Stephan C;  
XX: WPI; 1993-198976/25.

Endoplasmic reticu  
Endomembrane reten  
Endomembrane reten  
Endoplasmic reticu  
C-terminal sequenc  
C-terminal sequenc  
C-terminal fragmen  
Peptide used to mo  
Pseudomonas exotox  
KDEL peptide for t  
Organelle targetin  
Membrane retaining  
Organelle-targetin  
Endomembrane reten  
Endoplasmic reticu  
Pseudomonas sp. ex  
Pseudomonas sp. ex  
CD4+ T-cell activa  
Targeting and inte  
Endoplasmic reticu  
Amino acid sequenc  
Protein localisati  
Endoplasmic reticu  
Rat endoplasmic re  
Mammalian cytoplas  
Mammalian cytoplas  
Endoplasmic reticu  
Endoplasmic reticu  
Calreticulin endop  
Human ICAM-1 extra  
Amino acid sequenc  
ER lumen retention  
Protein localisati  
Peptide retention  
Amino acid sequenc  
C-terminal ER rete

XX Endoplasmic reticulum-located recombinant di:baeic endoprotease -  
 PT used in host cells to raise ratio of active, correctly folded  
 PT heterologous protein  
 XX  
 XX  
 PS Disclosure; Page 3; 30pp; English.  
 XX  
 CC Dibaic endoproteases of the invention are localised to the  
 CC endoplasmic reticulum of yeast host cells via ER-retention signals.  
 CC A preferred retention signal is the tetrapeptide KDEL functional in  
 CC mammalian cells. See also AAO42671, AAO42672 and AAR41971-2.  
 CC (Updated on 10-MAR-2003 to add missing OS field.)  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 CC (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 4 AA;  
 Query Match 100.0%; Score 20; DB 14; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 DB 1 KDEL 4  
 RESULT 2  
 AAR48246  
 ID AAR48246 standard; peptide; 4 AA.  
 AC AAR48246;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 29-JUL-1994 (first entry)  
 XX  
 DE Endoplasmic reticulum retention signal KDEL.  
 XX  
 KW Single chain antibody; sFv; heavy chain; light chain; kappa;  
 KW variable domain; hydrophilic linker; antibodies;  
 KW endoplasmic reticulum retention.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9402610-A1.  
 XX  
 PD 03-FEB-1994.  
 XX  
 PF 16-JUL-1993; 93WO-US06735.  
 XX  
 PR 17-JUL-1992; 92US-0916939.  
 PR 17-MAR-1993; 93US-0045274.  
 XX  
 PA (DAND ) DANA FARMER CANCER INST INC.  
 PI Haeeltine WA, Marasco WA;  
 XX  
 DR WPI; 1994-048868/06.  
 XX  
 PT Intracellular binding of antigens - by using antibody targeting  
 PT with vector system, for e.g. tumour suppression  
 XX  
 PS Claim 32; Page 99; 155pp; English.  
 XX  
 CC New vector systems comprise a sequence adapted for intracellular  
 CC delivery and expression contg. a promoter operably linked to an  
 CC antibody gene encoding an antibody which binds to a specific target  
 CC antigen. The antibody is esp. a single chain antibody in which the  
 CC heavy and light chain variable regions are joined via a hydrophilic  
 CC linker peptide. Localisation sequences are pref. included in the  
 CC constructs. See AAR48246-9 and AAR48252-3 for pref. (known) endoplasmic  
 CC reticulum retention signals.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX

SQ Sequence 4 AA;  
 Query Match 100.0%; Score 20; DB 15; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 DB 1 KDEL 4  
 RESULT 3  
 AAR49442  
 ID AAR49442 standard; Protein; 4 AA.  
 AC AAR49442;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 16-SEP-1994 (first entry)  
 XX  
 DE Immunomodulatory trafficking sequence #1.  
 XX  
 KW Naturally-occurring; immunomodulatory protein; human; therapy; class I;  
 KW major histocompatibility complex; class II; allotype; type I diabetes;  
 KW autoimmune disease; rheumatoid arthritis; T-cell-mediated response;  
 KW multiple sclerosis; transplant rejection; vaccine; MHC.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9404171-A1.  
 XX  
 PD 03-MAR-1994.  
 XX  
 PF 11-AUG-1993; 93WO-US07545.  
 XX  
 PR 11-AUG-1992; 92US-0925460.  
 PR 15-JUN-1993; 93US-0077255.  
 XX  
 PA (HARD ) HARVARD COLLEGE.  
 PI Chicz RM, Hedley ML, Stern LJ, Strominger JL, Urban RG;  
 PI Vignali DA;  
 XX  
 DR WPI; 1994-082825/10.  
 XX  
 PT Novel immunomodulatory peptide(s) and nucleic acids - useful for  
 PT treatment of auto-immune diseases, transplant rejection and for  
 PT vaccination  
 XX  
 PS Claim 13; Page 94; 139pp; English.  
 XX  
 CC The sequences given in AAR49291-505 and AAR46981-7038 represent peptide  
 CC fragments of naturally-occurring immunomodulatory proteins. These  
 CC fragments are between 10-30 residues in length and bind to a human  
 CC major histocompatibility complex (MHC) class II allotype. These  
 CC peptides may be used for therapy of autoimmune diseases, such as  
 CC type I diabetes, rheumatoid arthritis and multiple sclerosis, and to  
 CC reduce transplant rejection. They may also be used for vaccination  
 CC providing an exclusively T-cell-mediated response, which can be  
 CC class I or class-II based, or both, depending on the length and  
 CC character of the immunogenic peptides.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 XX  
 SQ Sequence 4 AA;  
 Query Match 100.0%; Score 20; DB 15; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 DB 1 KDEL 4

RESULT 4  
 AAR49584  
 ID AAR49584 standard; peptide; 4 AA.  
 XX  
 AC AAR49584;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 15-SEP-1994 (first entry)  
 XX  
 DE Sequence of endoplasmic reticulum (ER) retention peptide.  
 XX  
 KM Endoplasmic reticulum; ER; trafficking sequence.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9404557-A1.  
 XX  
 PD 03-MAR-1994.  
 XX  
 PF 11-AUG-1992; 92WO-US06692.  
 XX  
 PR 11-AUG-1992; 92WO-US06692.  
 XX  
 PA (HARD ) HARVARD COLLEGE.  
 XX  
 PI Chicz RM, Hedley ML, Stern LJ, Strominger JL, Urban RG;  
 PI Vignali DM;  
 XX  
 XX WPI; 1994-083102/10.  
 DR  
 PT New peptide binding to MHC class II allotype - useful for  
 PT treating autoimmune diseases, transplant rejection and for  
 PT immunisation  
 XX  
 PS Claim 20; Page 49; 60pp; English.  
 XX  
 CC A trafficking sequence is an AA sequence which functions to control  
 CC intracellular trafficking (directed movement from organelle to  
 CC organelle or to the cell surface) of a polypeptide to which it is  
 CC attached. Such trafficking sequences might traffic the polypeptide  
 CC to ER, a lysosome, or an endosome, and include signal peptides, ER  
 CC retention peptides such as AAR49584; and lysosome-targeting peptides  
 CC such as AAR4585 and AAR49586. An example of a signal peptide with such  
 CC a function is the signal peptide of MHC class II alpha (AAR49587).  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 CC  
 XX  
 SQ Sequence 4 AA;  
 XX  
 Query Match 100.0%; Score 20; DB 15; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 Db 1 KDEL 4  
 XX  
 RESULT 5  
 AAR95219  
 ID AAR95219 standard; protein; 4 AA.  
 XX  
 AC AAR95219;  
 XX  
 DT 16-DEC-1996 (first entry)  
 DT  
 XX  
 DE Pseudomonas exotoxin modified C-terminal sequence.  
 XX  
 KM Antibody; fusion protein; single chain; inhibition; tumour;  
 KM diagnosis; detection; imaging; immunotoxin; targeting; assay;  
 KM immunosassay; Lewis(Y) carbohydrate antigen.  
 XX

OS Pseudomonas sp.  
 XX  
 PN WO9613594-A1.  
 XX  
 PD 09-MAY-1996.  
 XX  
 PF 26-OCT-1995; 95WO-US13811.  
 XX  
 PR 28-OCT-1994; 94US-0311398.  
 PR 28-OCT-1994; 94US-0311398.  
 PR 28-OCT-1994; 94US-0311397.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Benhar I, Brinkmann U, Fitzgerald D, Jung S, Lee B;  
 PI Padlan EA, Pai L, Pastan I, Willingham M;  
 XX  
 XX WPI; 1996-251462/25.  
 DR  
 PT Single chain fusion proteins and antibodies - useful to diagnose and  
 PT treat cancer, specifically bind Lewis(Y) related carbohydrate  
 PT antigen  
 XX  
 PS Disclosure; Page 14; 116pp; English.  
 XX  
 CC A novel recombinant DNA molecule which encodes a single chain fusion  
 CC protein or antibody comprising the Fv region of both the light and  
 CC heavy chains of an antibody (Ab) fused together, and an effector  
 CC molecule, where the fusion protein or Ab has the binding specificity  
 CC of monoclonal Ab (MAb) B1, B3 or B5, can be used for the production  
 CC of such fusion proteins or antibodies. The fusion proteins can be  
 CC used in compositions as an immunotoxin to inhibit tumour cell growth.  
 CC The single chain antibody can be used to detect the presence or  
 CC absence of cells bearing a Lewis(Y) carbohydrate antigen in a  
 CC patient. The antibodies are also useful as multiple targeting  
 CC moieties, providing at least 2 kinds of biological activity. They  
 CC can also be used in diagnostic assays and for the imaging of tumours  
 CC when attached to a radiolabel and for the pathological diagnosis of  
 CC tumours. Humanised antibodies are less immunogenic than the mouse  
 CC MAbs B1, B3 and B5, making them more suitable for long term  
 CC treatment. The effector molecule used is preferably a drug or  
 CC cytotoxin which then produces an immunotoxin capable of selectively  
 CC killing particular cells. Preferred toxins are the Pseudomonas  
 CC exotoxin or Diphtheria toxin. These are both highly toxic compounds  
 CC and so are modified to eliminate domain Ia of the Pseudomonas toxin,  
 CC various amino acid deletions in domains II and III, single amino  
 CC acid substitutions and addition of one or more sequences at the  
 CC C-terminal end. The wild type C-terminal sequence of the Pseudomonas  
 CC exotoxin is given in AAR95221.  
 CC  
 XX  
 SQ Sequence 4 AA;  
 XX  
 Query Match 100.0%; Score 20; DB 17; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 Db 1 KDEL 4  
 XX  
 RESULT 6  
 AAR94944  
 ID AAR94944 standard; Protein; 4 AA.  
 XX  
 AC AAR94944;  
 XX  
 DT 31-OCT-1996 (first entry)  
 DT  
 XX  
 DE Microsomal retention signal.  
 XX  
 KM Toxin; subunit; vaccine; transgenic plant; immunogen; antigen;  
 KM adjuvant; immunisation.  
 XX

XX OS Synthetic.  
 XX PN WO9612801-A1.  
 XX PD 02-MAY-1996.  
 XX PF 24-OCT-1995; 95WO-US13376.  
 XX PR 24-OCT-1994; 94US-0328716.  
 XX PA (TULSA ) TULANE EDUCATIONAL FUND.  
 XX PA (TEXA ) UNTV TEXAS A & M SYSTEM.  
 XX PI Arntzen CJ, Clements JD, Haq TA, Mason HS;  
 XX DR WPI; 1996-230602/23.  
 XX PT Transgenic plants contg. E. coli heat labile enterotoxin subunits -  
 XX PS used as oral vaccines for animals which consume the plant  
 XX SQ Example 13; Page 95; 130pp; English.

CC A transgenic plant comprising or expressing a DNA sequence encoding  
 CC an immunogenic agent can be used as an oral vaccine for animals.  
 CC The vaccine is administered by the oral consumption of the plant and  
 CC provides the first known functional method for immunising animals  
 CC using transgenic plants, where the plants express bacterial antigens  
 CC that act as both immunogens and adjuvants. The method provides an  
 CC inexpensive production and delivery system for such antigens to  
 CC animals. This sequence acts as a microosomal retention signal and was  
 CC used to modify the LT-B Escherichia coli toxin subunit C-terminal  
 CC sequence. The sequence coding for the whole LT-B gene was used in the  
 CC construction of such a transgenic plant. The immunogenic agent  
 CC preferably comprises the LT-B or CT-B (cholera toxin B subunit) or  
 CC optionally LT-A or CT-A.

XX SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 17; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
 ||||  
 Db 1 KDEL 4

RESULT 7

AA95063  
 XX AAR95063 standard; peptide; 4 AA.

AC AAR95063;

DT 22-AUG-1996 (first entry)

DE Mammalian endoplasmic reticulum retention signal.

KW Nucleic acid transfer system; gene transfer; gene therapy;  
 KW cell targeting; multidomain protein; vector; cancer;  
 KW endoplasmic reticulum.

XX OS Synthetic.

PN WO9613599-A1.

PD 09-MAY-1996.

PF 31-OCT-1995; 95WO-EP04270.

PR 01-NOV-1994; 94EP-0810627.

PA (WEIS/) WEIS W.

XX XX Fominaya J, Weis W;  
 XX PI  
 XX DR WPI; 1996-239505/24.

PT Nucleic acid transfer system for gene therapy, e.g. against cancer  
 PT - includes toxin translocation domain to target nucleic acid to  
 PT specific cell

XX PS Disclosure; Page 11; 106pp; English.

CC Endoplasmic reticulum retention signals include a mammalian  
 CC signal (AAR95063), a bacterial signal (AAT05064) from Pseudomonas  
 CC exotoxin, and a yeast signal (AAR95065). The signal may form  
 CC part of a multidomain protein (see also AAR95053-58) that is used  
 CC with an effector nucleic acid for the transfer of nucleic acids  
 CC to targeted cells as a means of gene therapy. The endoplasmic  
 CC reticulum retention signal functions to affect intracellular  
 CC routing of the internalized protein/nucleic acid complex.

XX SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 17; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
 ||||  
 Db 1 KDEL 4

RESULT 8  
 AAM31861  
 ID AAM31861 standard; peptide; 4 AA.

XX AC AAM31861;

DT 06-MAY-1998 (first entry)

DE Endoplasmic reticulum retention peptide.

KW RNA-loaded antigen presenting cell; trafficking sequence; APC production;  
 KW tumour formation; pathogen infection; antigenic epitope; immune response;  
 KW T cell proliferation; cytotoxic T lymphocyte; adoptive immunotherapy;  
 KW therapy; TAB; CTL; PAB; endoplasmic reticulum retention peptide.

XX OS Synthetic.

PN WO9741210-A1.

PD 06-NOV-1997.

PF 30-APR-1997; 97WO-US07317.

PR 30-APR-1996; 96US-0640444.

PA (UYDU-) UNIV DUKE.

PI Boczkowski DJ, Gilboa E, Nair SK;

DR WPI; 1997-549715/50.

PT Use of RNA-loaded antigen presenting cells - to induce T-cell  
 PT proliferation for the treatment or prevention of tumour formation or  
 PT pathogen infection

PS Claim 49; Page 38; 47pp; English.

CC This sequence represents a endoplasmic reticulum retention peptide, and  
 CC can be used in the method of the invention. The method is for producing  
 CC an RNA-loaded antigen presenting cell (APC) that presents on its surface  
 CC a tumour or pathogen antigenic epitope (TAB or PAB respectively) that  
 CC induces T cell proliferation and an immune response against the tumour or



CC pathogen, and comprises introducing into an APC in vitro, RNA that  
CC encodes the antigen. The RNA-loaded APCs can be used to stimulate  
CC cytotoxic T lymphocyte (CTL) proliferation ex vivo or in vivo. The ex  
CC vivo expanded CTL can be administered to a patient in a method of  
CC adoptive immunotherapy. The methods can be used for treating or  
CC preventing tumour formation or pathogen infection caused by e.g. HIV,  
CC hepatitis, influenza, poliomyelitis, measles, herpes, mumps or rubella  
CC viruses, Salmonella, Shigella or Enterobacter. The method circumvents the  
CC need to purify RNA or isolate and identify a TAE or PAE.

SO Sequence 4 AA;

Query Match 100.0%; Score 20; DB 18; Length 4;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
1 KDEL 4

Db

RESULT 9  
AAW21969 standard; peptide; 4 AA.

AAW21969;

03-DEC-1997 (first entry)

Linker #3 for immunotoxin containing Pseudomonas exotoxin.

PCR; primer; amplify; polymerase chain reaction; antibody; immunotoxin;  
variable heavy chain; VH; murine monoclonal antibody; Lewis; carcinoma;  
carbohydrate antigen; Pseudomonas exotoxin; proteolytic activation;  
cytotoxic activity; tumour; autoimmune condition; rheumatoid arthritis;  
graft versus host disease; organ transplant rejection; type I diabetes;  
multiple sclerosis; systemic lupus erythematosus; myasthenia gravis;  
T cell; B cell; cytosol; bone marrow; transplant; therapy.

Synthetic.

MO9713529-A1.

17-APR-1997.

11-OCT-1996; 96MO-US16327.

13-OCT-1995; 95US-0005388.

(USSH ) US DEPT HEALTH & HUMAN SERVICES.

Kuan C, Pastan I;

WPI; 1997-235666/21.

Immunotoxin(s) comprising Pseudomonas exotoxin linked to  
disulphide stabilised variable heavy and light chain regions of an  
antibody - useful for killing target cells bearing characteristic  
marker

Claim 9; Page 50; 64pp; English.

AAW21967-W21969 represent linkers used in the immunotoxins of the  
invention. The immunotoxins bind to target cells, and comprise, a  
Pseudomonas exotoxin (PE) that does not need proteolytic activation for  
cytotoxic activity fused to a VH framework region of an Fv antibody (Ab)  
fragment. The VH chain region is bound through at least one disulphide  
bond to a variable light (VL) chain framework region. The PE is lacking  
residues 1-279 and is at least 10-fold more cytotoxic to the target  
cells than an immunotoxin comprising PE attached to a VH chain framework  
region of an Fv Ab fragment lacking a disulphide bond to a VL chain  
framework region. These sequences are used to join the VH chain region to  
the PE. The immunotoxins can be used for killing target cells in the

CC treatment of tumours, autoimmune conditions, graft versus host disease,  
CC organ transplant rejection, type I diabetes, multiple sclerosis,  
CC rheumatoid arthritis, systemic lupus erythematosus, myasthenia gravis,  
CC etc, all caused by T and B cells. They can also be used to deliver an  
CC antibody to the cytosol of a cell, and in vitro in the elimination of  
CC haematil cells from bone marrow before transplant. The immunotoxins have  
CC high cytotoxicity to target cells and a small size to provide greater  
CC penetration to target cells.

SO Sequence 4 AA;

Query Match 100.0%; Score 20; DB 18; Length 4;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
1 KDEL 4

Db

RESULT 10  
AAW76396 standard; Peptide; 4 AA.

AAW76396;

11-JAN-1999 (first entry)

Endoplasmic reticulum transport signal consensus.

Exotoxin A; ETA; drug delivery; membrane transport;  
endoplasmic reticulum.

Synthetic.

MO9842876-A1.

01-OCT-1998.

24-MAR-1998; 98MO-US05710.

26-MAR-1997; 97US-0042056.

(TEXA ) UNIV TEXAS SYSTEM.

Draper RK;

WPI; 1998-532023/45.

Delivering compounds to cells as new conjugate with detoxified  
exotoxin A - able to cross membranes and deliver to the cytoplasm,  
e.g. nucleic acids, antibodies, tumour suppressors etc.

Disclosure; Page 8; 76pp; English.

This peptide is the consensus intracellular transport signal used  
for returning back to the endoplasmic reticulum (ER) proteins that  
have escaped the ER and entered to Golgi complex. The mechanism  
involves the KDEL receptor. The KDEL motif can replace a REDL  
motif (see AAW76395) present at the C-terminal end of Pseudomonas  
aeruginosa exotoxin A (ETA, see AAW76391). This motif is important  
in the intracellular transport and cytotoxicity of ETA. This  
suggests that ETA interacts with the KDEL receptor and that ETA  
may reach the interior of the ER before penetrating the cytosol.  
The invention provides a means of delivering compounds to cells as  
conjugates with modified ETA. The modified ETA is able to cross  
membranes and deliver e.g. therapeutic agents to the cytoplasm,  
such as nucleic acids, peptides, peptide nucleic acids, single  
chain antibodies and tumour suppressors.

Sequence 4 AA;

Query Match 100.0%; Score 20; DB 19; Length 4;

Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
DB 1 KDEL 4

## RESULT 11

AAW64614 standard; Protein; 4 AA.

AC AAW64614;

DT 03-NOV-1998 (first entry)

DE Endomembrane retention signal peptide.

XX Target; imaging agent; epithelium; transepithelial transport; diagnosis;

KW transcytosis; disease; basolateral; internalisation; J chain.

XX Unknown.

PN MO9830591-A1.

PD 16-JUL-1998.

PF 09-JAN-1998; 98WO-US00339.

PR 10-JAN-1997; 97US-0782480.

PA (EPIC-) EPICYTE PHARM INC.

PI Fitchen JH, Hein MB, Hiatt AC;

DR WPI; 1998-399066/34.

PT New epithelial tissue targeting agent - used to deliver imaging  
agents to an epithelial surface for internalisation; useful in  
diagnosis

PS Example 1c; Page 90; 118pp; English.

CC This sequence represents an endomembrane retention signal which is used  
in a method involving the construction of a target molecule from human J  
chain protein fragments. This construct is used in a method to target  
imaging agents to epithelial surfaces at which they may remain or undergo  
transepithelial transport via transcytosis. At least one imaging agent is  
linked to the targeting molecule comprising a polypeptide that (a) forms  
a closed covalent loop, (b) contains at least 3, preferably 4, peptide  
domains having beta-sheet character separated by domains lacking  
beta-sheet character and (c) is not full length dimeric IGA. The imaging  
agents are useful in the diagnosis of disease. The target molecule is  
also capable of specifically binding to a basolateral factor associated  
with an epithelial surface to cause internalisation of a biological agent  
linked to the target molecule.

SO Sequence 4 AA;

Query Match 100.0%; Score 20; DB 19; Length 4;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
DB 1 KDEL 4

## RESULT 12

AAW61591 standard; Peptide; 4 AA.

AC AAW61591;

XX 27-OCT-1998 (first entry)  
DT Endomembrane retention signal.  
XX

KW J chain; targeting molecule; epithelial; beta-sheet; asthma;  
cancer; inflammatory disorder; autoimmune disorder; celiac disease;  
colitis; pneumonia; cystic fibrosis.

XX Synthetic.

PN MO9830592-A1.

PD 16-JUL-1998.

PF 09-JAN-1998; 98WO-US00542.

PR 10-JAN-1997; 97US-0782481.

PA (EPIC-) EPICYTE PHARM INC.

PI Fitchen JH, Hein MB, Hiatt AC;

DR WPI; 1998-399067/34.

PT New epithelial tissue targeting agent - used to deliver  
biologically active compounds to an epithelial surface for  
internalisation

PS Example 1; Page 48; 142pp; English.

CC The endomembrane retention signal is used in the synthesis of a targeting  
molecule (TM). The TMs are used to target biological agents to  
epithelial surfaces at which they can be internalised. The TMs comprise  
a polypeptide that: (a) forms a closed covalent loop; (b) contains at  
least 3, preferably 4, peptide domains having beta-sheet character  
separated by domains lacking beta-sheet character; and (c) is not full  
length dimeric IGA. The TMs are useful to prevent and/or treat diseases  
associated with epithelial surfaces, e.g. asthma, cancer,  
(myco)bacterial, viral or fungal infection, inflammatory disorders,  
autoimmune disorders, celiac disease, colitis, pneumonia and cystic  
fibrosis.

SO Sequence 4 AA;

Query Match 100.0%; Score 20; DB 19; Length 4;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
DB 1 KDEL 4

## RESULT 13

AAW56513 standard; Peptide; 4 AA.

AC AAW56513;

DT 28-SEP-1998 (first entry)

DE Endoplasmic reticulum recycling peptide.

KW Intracellular targeting sequence; endoplasmic reticulum;

KW DNA vaccine; genetic immunisation; allergy; autoimmune disease;

XX cancer; infection; psoriasis.

OS Synthetic.

PN MO9817323-A1.

PD 30-APR-1998.

XX 23-OCT-1997; 97WO-US19545.  
 XX 23-OCT-1996; 96US-0029592.  
 XX (TYPE-) UNIV PENNSYLVANIA.  
 XX Madala M, Weiner DB, Williams WV;  
 XX WPI; 1998-261196/23.  
 XX  
 PT Plasmid encoding immunogenic target protein - used in, e.g.  
 PT protective or therapeutic vaccines against allergy, cancer,  
 PT microbial infection or autoimmune disease  
 PS Claim 7; Page 62; 84pp; English.  
 XX  
 CC This peptide, when joined to the C-terminus of a protein, acts as  
 CC an endoplasmic reticulum (ER) recycling signal that localises the  
 CC protein from the Golgi body back to the ER when expressed in a  
 CC cell. A claimed novel plasmid comprises a sequence, linked to  
 CC regulatory elements, that encodes an immunogenic target protein  
 CC that includes, or is linked to, an intracellular targeting sequence  
 CC (ITS) such as this peptide. Other ITS peptides (see AAM56512 and  
 CC AAM56514) are provided that direct localisation in the lysosome or  
 CC the ER. The novel plasmid is used as a protective or  
 CC therapeutic DNA vaccine to immunise against the immunogenic target  
 CC protein (claimed), particularly in cases of allergy, cancer (or  
 CC other proliferative diseases such as psoriasis), microbial  
 CC infection or autoimmune disease, e.g. rheumatoid arthritis,  
 CC insulin-dependent diabetes mellitus, Crohn's disease, asthma,  
 CC penicillin anaemia and many others. A particular use is against  
 CC pathogens that, at least for part of their life cycle, are  
 CC intracellular, e.g. hepatitis C, HIV, Neisseria gonorrhoeae,  
 CC listeria and Shigella. The cytotoxic T cell response is enhanced  
 CC by delivering the immunogenic target protein to the ER.  
 CC  
 SQ Sequence 4 AA;  
 Query Match 100.0%; Score 20; DB 19; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 DB 1 KDEL 4  
 RESULT 14  
 AAM51434  
 ID AAM51434 standard; peptide; 4 AA.  
 AC AAM51434;  
 XX  
 DT 02-SEP-1998 (first entry)  
 XX  
 DE C-terminal sequence of localisation peptides for the ER.  
 XX  
 KW Bifunctional chelating polysaccharide; targeting peptide; heparin;  
 KW nuclear localisation; fluorescent indicator; dextran; ficol; glycogen;  
 KW amylopectin; mannan; inulin; starch; agarose; cellulose; ER;  
 KW endoplasmic reticulum.  
 XX  
 OS Synthetic.  
 XX  
 XX US573227-A.  
 XX  
 PD 30-JUN-1998.  
 XX  
 PF 23-JUN-1993; 93US-0082269.  
 XX  
 PR 23-JUN-1993; 93US-0082269.  
 XX

PA (MOLE-) MOLECULAR PROBES INC.  
 XX  
 PI Albritton NL, Kuhn MA, Meyer T;  
 XX  
 DR WPI; 1998-386995/33.  
 XX  
 PT Bifunctional chelating polysaccharides used for analysing  
 PT intracellular ion levels - have chelating moieties attached which  
 PT act as a fluorescent indicator in the presence of selected metal  
 PT ions  
 XX  
 PS Disclosure; Column 9; 29pp; English.  
 XX  
 CC The invention relates to a water soluble polysaccharide with a molecular  
 CC weight greater than 1000 Daltons covalently attached to (a) at least one  
 CC chelating moiety selective for a monovalent or divalent metal ion (to  
 CC act as a fluorescent indicator for the metal ion) and (b) at least one  
 CC targeting peptide (to localise the compound to the inside of a cellular  
 CC organelle). Preferably the polysaccharide is dextran, ficol, heparin,  
 CC glycogen, amylopectin, mannan, inulin, starch, agarose or cellulose with  
 CC a molecular weight less than 3,000,000. The metal ion is preferably Na<sup>+</sup>,  
 CC K<sup>+</sup>, Li<sup>+</sup>, Ca<sup>2+</sup> or Mg<sup>2+</sup>. The compounds are useful for analysing intra-  
 CC cellular ion levels, especially Ca<sup>2+</sup>. The present sequence represents a  
 CC carboxyl terminal sequence commonly shared by localisation peptides  
 CC for the endoplasmic reticulum (ER).  
 CC  
 SQ Sequence 4 AA;  
 Query Match 100.0%; Score 20; DB 19; Length 4;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 DB 1 KDEL 4  
 RESULT 15  
 AAM51437  
 ID AAM51437 standard; peptide; 4 AA.  
 AC AAM51437;  
 XX  
 DT 02-SEP-1998 (first entry)  
 XX  
 DE C-terminal sequence of endoplasmic reticulum retained proteins.  
 XX  
 KW Endoplasmic reticulum; carboxy terminus; Golgi apparatus;  
 KW glycosyltransferase; membrane anchor; oligosaccharide synthesis.  
 XX  
 OS Synthetic.  
 XX  
 XX US5776772-A.  
 XX  
 PD 07-JUL-1998.  
 XX  
 PF 30-JAN-1996; 96US-0593865.  
 XX  
 PR 27-MAY-1992; 92US-0849045.  
 PR 24-OCT-1989; 89US-0426577.  
 PR 10-MAR-1994; 94US-0209604.  
 PR 30-JAN-1996; 96US-0593865.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Adler B, Browne JK, Colley KJ, Paulson JG, Ujita-Jee E,  
 PI Weinstein J;  
 XX  
 DR WPI; 1998-398046/34.  
 XX  
 PT Expression cassette for producing glycosyltransferase in secretable  
 PT form - lacking membrane anchor and Golgi retention signal, used for  
 PT synthesis of oligosaccharide(s)

XX Disclosure; Column 3; 16pp; English.

PS  
XX  
CC The invention relates to expression cassettes which contain a promoter  
CC and a DNA sequence encoding a secreted glycosyltransferase that lacks  
CC the membrane anchor and the Golgi retention signal. The glycosyl-  
CC transferases are used in enzymatic oligosaccharide synthesis.  
CC They are produced in a soluble form that is secreted into the culture  
CC medium without loss of enzymatic activity, resulting in increased  
CC production and simpler recovery (free of any membrane-bound glycosyl-  
CC transferase), compared with extraction from mammalian tissue. The  
CC present sequence represents a carboxy-terminal sequence sufficient for  
CC retention of a protein in the endoplasmic reticulum.

XX  
SQ Sequence 4 AA;

Query Match 100.0%; Score 20; DB 19; Length 4;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 1 KDEL 4

Search completed: August 26, 2003, 16:04:29  
Job time : 11.2832 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: August 26, 2003, 16:05:48 ; Search time 1.41593 Seconds  
(without alignments)  
119.528 Million cell updates/sec

Title: US-09-696-872-37

Perfect score: 20

Sequence: 1 KDEL 4

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PTCTUS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	4	1	US-08-328-961-6
2	20	100.0	4	1	US-07-872-673B-5
3	20	100.0	4	1	US-08-405-615-16
4	20	100.0	4	1	US-08-331-398A-51
5	20	100.0	4	1	US-08-462-397-6
6	20	100.0	4	1	US-08-270-314-3
7	20	100.0	4	1	US-08-406-192-28
8	20	100.0	4	1	US-08-082-269D-7
9	20	100.0	4	1	US-08-593-865-1
10	20	100.0	4	2	US-08-461-234-16
11	20	100.0	4	2	US-08-480-190-152
12	20	100.0	4	2	US-08-545-151-28
13	20	100.0	4	2	US-08-373-190-17
14	20	100.0	4	2	US-08-463-480-16
15	20	100.0	4	2	US-08-470-566B-51
16	20	100.0	4	2	US-08-488-379-152
17	20	100.0	4	2	US-08-821-840-5
18	20	100.0	4	2	US-08-713-928B-15
19	20	100.0	4	2	US-08-407-900B-6
20	20	100.0	4	2	US-08-419-075-28
21	20	100.0	4	2	US-08-438-190A-17
22	20	100.0	4	2	US-08-923-536A-2
23	20	100.0	4	2	US-08-491-988-29
24	20	100.0	4	2	US-08-665-202-29
25	20	100.0	4	2	US-08-809-668-9
26	20	100.0	4	2	US-08-331-397B-51
27	20	100.0	4	2	US-08-759-804A-51

28	20	100.0	4	2	US-08-818-253-49	Sequence 49, Appl
29	20	100.0	4	3	US-08-350-215-17	Sequence 17, Appl
30	20	100.0	4	3	US-08-722-258-62	Sequence 62, Appl
31	20	100.0	4	3	US-08-718-904-42	Sequence 42, Appl
32	20	100.0	4	3	US-08-782-480-44	Sequence 44, Appl
33	20	100.0	4	3	US-09-287-145A-17	Sequence 17, Appl
34	20	100.0	4	3	US-09-397-951-9	Sequence 9, Appl
35	20	100.0	4	3	US-08-776-271-7	Sequence 7, Appl
36	20	100.0	4	3	US-09-047-148-12	Sequence 12, Appl
37	20	100.0	4	3	US-09-094-359-17	Sequence 17, Appl
38	20	100.0	4	3	US-09-172-063-37	Sequence 37, Appl
39	20	100.0	4	3	US-08-789-333F-28	Sequence 28, Appl
40	20	100.0	4	3	US-09-215-035-7	Sequence 7, Appl
41	20	100.0	4	3	US-09-124-671-37	Sequence 37, Appl
42	20	100.0	4	3	US-09-169-015-38	Sequence 38, Appl
43	20	100.0	4	3	US-09-169-425C-23	Sequence 23, Appl
44	20	100.0	4	3	US-09-191-852-1	Sequence 1, Appl
45	20	100.0	4	3	US-08-818-252-49	Sequence 49, Appl

## ALIGNMENTS

RESULT 1  
US-08-328-961-6  
; Sequence 6, Application US/08328961  
; Patent No. 5501975  
; GENERAL INFORMATION:  
; APPLICANT: Chaudhuri, Bhabatosh  
; APPLICANT: Stephan, Christine  
; APPLICANT: Seeboth, Peter  
; APPLICANT: Reizman, Howard  
; TITLE OF INVENTION: No. 5501975el DNA Molecules and Hosts  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Ciba-Geigy Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/328,961  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/989,260  
; FILING DATE: 11-DEC-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spruill, W. Murray  
; REGISTRATION NUMBER: 32,943  
; REFERENCE/DOCKET NUMBER: 4-18885/A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8614  
; TELEFAX: 919-541-8669  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FEATURE:  
; NAME/KEY: Domain  
; LOCATION: 1..4  
; OTHER INFORMATION: /note= "ER retention signal KDEL"  
US-08-328-961-6  
Query Match 100.0%; Score 20; DB 1; Length 4;

Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
|||  
Db 1 KDEL 4

## RESULT 2

US-07-872-673B-5  
; Sequence 5, Application US/07872673B  
; Patent No. 5578466

; GENERAL INFORMATION:

APPLICANT: Toshiya HAYANO, Setsuko KATO, No. 5578466hiro TAKAHASHI, and Masanori  
TITLE OF INVENTION: Co-expression System of Protein Disulfide Isomerase Gene an

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS & CUSHMAN  
STREET: 130 WATER STREET

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Type 2DD, 3.50 inch, 720 KB

COMPUTER: Apple Macintosh SE

OPERATING SYSTEM: Apple DOS

SOFTWARE: Microsoft Word Version 4.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/872,673B

FILING DATE: 19920417

CLASSIFICATION: 435

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER: Japanese Patent Application No. 5578466, 114074/91 and 31160

FILING DATE: 18-APR-1991 and 30-OCT-1991

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 4

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: C-terminal fragment

FEATURE:

NAME/KEY: ER retention signal

LOCATION: C-terminus

IDENTIFICATION METHOD: ER retention of proteins having this signal

OTHER INFORMATION: located at the C-terminus of rat PDI

PUBLICATION INFORMATION:

AUTHORS: Edman, Jeffrey C., Ellis, Ireland, Blacher, Russell W., Roth, Richa

TITLE: Sequence of protein disulfide isomerase and implications of its rel

PATENT No. 5578466

JOURNAL: Nature

VOLUME: 317

PAGES: 267-270

DATE: 19-Sep-1985

RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 486 to 489

US-07-872-673B-5

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
|||  
Db 1 KDEL 4

RESULT 3  
US-08-405-615-16  
; Sequence 16, Application US/08405615  
; Patent No. 5602095  
; GENERAL INFORMATION:

APPLICANT: Pastan, Ira  
APPLICANT: Fitzgerald, David J.  
TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with  
TITLE OF INVENTION: Increased Activity

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ellen L. Weber  
STREET: One Market Plaza, Stewart Tower, Suite 2000

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/405,615

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/901,709

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Ellen L.

REGISTRATION NUMBER: 32,762

REFERENCE/DOCKET NUMBER: 15280-36

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-543-9600

TELEFAX: 415-543-5043

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 4 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-405-615-16  
Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
|||  
Db 1 KDEL 4

RESULT 4  
US-08-331-398A-51  
; Sequence 51, Application US/08331398A  
; Patent No. 5608039

; GENERAL INFORMATION:

APPLICANT: Pastan, Ira

APPLICANT: Williamson, Mark

APPLICANT: Fitzgerald, David

APPLICANT: Brinkmann, Ulrich

APPLICANT: Pai, Lee

TITLE OF INVENTION: Single Chain B3 Antibody Fusion Proteins

TITLE OF INVENTION: and Their Uses (as amended)

NUMBER OF SEQUENCES: 68

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew

STREET: One Market Plaza, Stewart Street Plaza

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94105-1492

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/331.398A  
FILING DATE: 28-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/767,331  
FILING DATE: 30-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/596,289  
FILING DATE: 12-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Hunter, Tom  
REGISTRATION NUMBER: 38,498  
REFERENCE/DOCKET NUMBER: 015280-126110US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-331-398A-51

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
Db 1 KDEL 4

RESULT 5  
US-08-462-397-6  
Sequence 6, Application US/08462397  
Patent No. 5618690  
GENERAL INFORMATION:  
APPLICANT: Chaudhuri, Bhahatosh  
APPLICANT: Stephan, Christine  
APPLICANT: Seebach, Peter  
APPLICANT: Reizman, Howard  
TITLE OF INVENTION: No. 5618690e1 DNA Molecules and Hosts  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESS: Ciba-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10512  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,397  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/989,260  
FILING DATE: 11-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Spruill, W. Murray  
REGISTRATION NUMBER: 32,943  
REFERENCE/DOCKET NUMBER: 4-18885/A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8614  
TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Domain  
LOCATION: 1..4  
OTHER INFORMATION: /note= "ER retention signal KDEL"  
US-08-462-397-6

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
Db 1 KDEL 4

RESULT 6  
US-08-270-314-3  
Sequence 3, Application US/08270314  
Patent No. 568388  
GENERAL INFORMATION:  
APPLICANT: CAMPBELL, Anthony K.  
TITLE OF INVENTION: MODIFIED BIOLUMINESCENT PROTEINS AND  
THEIR USE  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: VA  
COUNTRY: US  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/270,314  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 89 16806.6  
FILING DATE: 22-JUL-1989  
APPLICATION NUMBER: PCT/GB90/01131  
FILING DATE: 23-JUL-1990  
APPLICATION NUMBER: US 07/820,867  
FILING DATE: 22-JAN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 521-2297  
TELEFAX: (703) 685-0573  
TELEX: 248425 EMBON  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-270-314-3

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4

Db 1 KDEL 4

## RESULT 7

US-08-406-192-28  
; Sequence 28, Application US/08406192  
; Patent No. 5739287  
; GENERAL INFORMATION:  
; APPLICANT: Wilbur, D. Scott  
; APPLICANT: Prathare, Pradip M  
; TITLE OF INVENTION: Biotinylated Cobalamins  
; NUMBER OF SEQUENCES: 47  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 1420 Fifth Avenue, Suite 2800  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: U.S.A.  
; ZIP: WA 98101-2333  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/406,192  
; FILING DATE: 16-MAR-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/224,831  
; FILING DATE: 08-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Renzoni, George E  
; REGISTRATION NUMBER: 37,919  
; REFERENCE/DOCKET NUMBER: RECL18947  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 682 8100  
; TELEFAX: (206) 224 0779  
; TELEX: 4938023  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-406-192-28

QY 1 KDEL 4  
Db 1 KDEL 4

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## RESULT 8

US-08-082-269D-7  
; Sequence 7, Application US/08082269D  
; Patent No. 5773227  
; GENERAL INFORMATION:  
; APPLICANT: Kuhn, Michael  
; APPLICANT: Meyer, Tobias  
; APPLICANT: Albritton, Nancy  
; TITLE OF INVENTION: Bifunctional Chelating Polysaccharides  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Molecular Probes, Inc.  
; STREET: 4849 Pitchford Avenue  
; CITY: Eugene  
; STATE: Oregon

COUNTRY: USA  
ZIP: 97402-9144  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch  
COMPUTER: IBM  
OPERATING SYSTEM: MS-DOS 6.2  
SOFTWARE: Text Editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/082,269D  
FILING DATE: 23-June-1993

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Hellesen, Ailegra J.  
REGISTRATION NUMBER: 34,179  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (503)465-8300  
TELEFAX: (503)344-6504  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 AMINO ACIDS  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
; MOLECULE TYPE: Peptide  
; HYPOTHETICAL: no  
; FRAGMENT TYPE:  
; PUBLICATION INFORMATION:  
; AUTHORS: Munro, Sean and Pelham Hugh, R.B.  
; TITLE: A C-Terminal Signal Prevents Secretion of Luminal ER Proteins  
; JOURNAL: Cell  
; VOLUME: 48  
; ISSUE: 13 March 1987  
; PAGES: 899-907  
; DATE: 1987  
US-08-082-269D-7

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
Db 1 KDEL 4

## RESULT 9

US-08-593-865-1  
; Sequence 1, Application US/08593865  
; Patent No. 5776772  
; GENERAL INFORMATION:  
; APPLICANT: Paulson, James C.  
; APPLICANT: Ujita-Lee, Bryn  
; APPLICANT: Colley, Karen J.  
; APPLICANT: Adler, Beverly  
; APPLICANT: Browne, Jeffrey K.  
; TITLE OF INVENTION: Method for Producing Secretable  
; TITLE OF INVENTION: Glycosyltransferases and Other Golgi Processing Enzymes  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Poma, Smith, Lande & Rose  
; STREET: 2029 Century Park East, 38th Floor  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90067  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/593,865  
; FILING DATE: January 30, 1996



CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO 91/06635  
FILING DATE: 16 May 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Oidenkamp, David J.  
REGISTRATION NUMBER: 29,421  
REFERENCE/DOCKET NUMBER: 117-033  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (310)788-5000  
TELEFAX: (310) 277-1297  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-593-865-1

Query Match 100.0%; Score 20; DB 1; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4  
Db 1 KDEL 4

RESULT 10  
US-08-461-234-16  
Sequence 16, Application US/08461234  
Patent No. 5821238

GENERAL INFORMATION:  
APPLICANT: Pasten, Ira H.  
APPLICANT: Fitzgerald, David J.  
TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with  
TITLE OF INVENTION: Increased Activity  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stuart Street Tower  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/461,234  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/901,709  
FILING DATE: 18-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,615  
FILING DATE: 15-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen Lauver  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 15280-36-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-08-461-234-16

Query Match 100.0%; Score 20; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4  
Db 1 KDEL 4

RESULT 11  
US-08-480-190-152  
Sequence 152, Application US/08480190  
Patent No. 5827516

GENERAL INFORMATION:  
APPLICANT: Robert G. Urban  
APPLICANT: Roman M. Chic  
APPLICANT: Dario A. A. Vignali  
APPLICANT: Mary L. Hedley  
APPLICANT: Lawrence J. Stern  
APPLICANT: Jack L. Strominger  
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES  
NUMBER OF SEQUENCES: 274  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S. A.  
ZIP: 02110-2804

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,190  
FILING DATE:

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/077,255  
FILING DATE: June 15, 1993  
APPLICATION NUMBER: 07/925,460  
FILING DATE: August 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00246/168001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 152:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-480-190-152

Query Match 100.0%; Score 20; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4  
Db 1 KDEL 4

RESULT 12

US-08-545-151-28  
Sequence 28, Application US/08545151  
Patent No. 5840712  
GENERAL INFORMATION:  
APPLICANT: Morgan Jr., A. Charles  
APPLICANT: Wilbur, D. Scott  
TITLE OF INVENTION: Water Soluble Vitamin B12 Receptor  
TITLE OF INVENTION: Modulating Agents and Methods Related Thereto  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson & KindnessP LLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: Washington  
COUNTRY: U.S.A.  
ZIP: WA 98101-2333  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,151  
FILING DATE: 19-OCT-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04404  
FILING DATE: 07-APR-1995  
APPLICATION NUMBER: US 08/406,192  
FILING DATE: 16-MAR-1995  
APPLICATION NUMBER: US 08/406,194  
FILING DATE: 16-MAR-1995  
APPLICATION NUMBER: US 08/406,191  
FILING DATE: 16-MAR-1995  
APPLICATION NUMBER: US 08/224,831  
FILING DATE: 08-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Renzoni, George E  
REGISTRATION NUMBER: 37,919  
REFERENCE/DOCKET NUMBER: RECL18878  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 224 0779  
TELEFAX: (206) 682 8100  
TELEX: 4938023  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-545-151-28

Query Match 100.0%; Score 20; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
Db 1 KDEL 4

RESULT 13  
US-08-373-190-17  
Sequence 17, Application US/08373190  
Patent No. 5851829  
GENERAL INFORMATION:  
APPLICANT: MARASCO, WAYNE  
APPLICANT: HASELTINE, WILLIAM  
TITLE OF INVENTION: METHOD OF INTRACELLULAR BINDING OF PROTEINS  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN  
STREET: 130 WATER STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Pasteo Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/373,190  
FILING DATE: 17-JAN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/06735  
FILING DATE: 16-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, DAVID S  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 41956-PCT-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX: STRE UR 2002  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
US-08-373-190-17

Query Match 100.0%; Score 20; DB 2; Length 4;  
Best Local Similarity 100.0%; Pred. No. 2.5e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
Db 1 KDEL 4

RESULT 14  
US-08-463-480-16  
Sequence 16, Application US/08463480  
Patent No. 5854044  
GENERAL INFORMATION:  
APPLICANT: Pastan, Ira H.  
APPLICANT: Fitzgerald, David J.  
TITLE OF INVENTION: Recombinant Pseudomonas Exotoxin with  
TITLE OF INVENTION: Increased Activity  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/463,480  
FILING DATE: 05-JUN-1995

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/901,709  
FILING DATE: 18-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,615  
FILING DATE: 15-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Weber, Ellen Lauver  
REGISTRATION NUMBER: 32,762  
REFERENCE/DOCKET NUMBER: 15280-36-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-08-463-480-16

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 4;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
Db 1 KDEL 4

RESULT 15  
US-08-470-566B-51  
Sequence 51, Application US/08470566B  
Patent No. 5872212  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Kozziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kriesty  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: No. 5872212e1 Pesticidal Proteins and Strains  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 5872212artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,566B  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 05-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018

FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Weigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-08-470-566B-51

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 4;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
Db 1 KDEL 4

Search completed: August 26, 2003, 16:19:16  
Job time : 2.41593 secs

This Page Blank (unspdo)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: August 26, 2003, 16:02:18 ; Search time 1.41593 Seconds  
(without alignments)  
372.792 Million cell updates/sec

Title: US-09-696-872-37  
Perfect score: 20  
Sequence: 1 KDEL 4

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 497079 seqs, 131961718 residues

Total number of hits satisfying chosen parameters: 497079

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications\_AA:\*  
2: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	4	8 US-08-765-244-16	Sequence 16, Appl
2	20	100.0	4	8 US-08-873-601-24	Sequence 24, Appl
3	20	100.0	4	9 US-09-157-748-30	Sequence 30, Appl
4	20	100.0	4	9 US-09-759-960-23	Sequence 23, Appl
5	20	100.0	4	9 US-09-854-122-3	Sequence 3, Appl
6	20	100.0	4	9 US-09-347-064-35	Sequence 35, Appl
7	20	100.0	4	9 US-09-333-527-2	Sequence 2, Appl
8	20	100.0	4	9 US-09-785-921A-16	Sequence 16, Appl
9	20	100.0	4	10 US-09-997-956-2	Sequence 2, Appl
10	20	100.0	4	10 US-09-781-804-2	Sequence 28, Appl
11	20	100.0	4	10 US-09-916-940-28	Sequence 8, Appl
12	20	100.0	4	10 US-09-480-236-8	Sequence 3, Appl
13	20	100.0	4	10 US-09-984-183-3	Sequence 22, Appl
14	20	100.0	4	10 US-09-999-745-22	Sequence 10, Appl
15	20	100.0	4	10 US-09-967-772-10	Sequence 10, Appl

16	20	100.0	4	10 US-09-554-000-49	Sequence 49, Appl
17	20	100.0	4	10 US-09-178-286-19	Sequence 19, Appl
18	20	100.0	4	10 US-09-792-630-77	Sequence 77, Appl
19	20	100.0	4	10 US-09-915-789-10	Sequence 10, Appl
20	20	100.0	4	10 US-09-807-721-6	Sequence 6, Appl
21	20	100.0	4	10 US-09-825-803-3	Sequence 3, Appl
22	20	100.0	4	11 US-09-575-847-8	Sequence 8, Appl
23	20	100.0	4	11 US-09-991-209-97	Sequence 97, Appl
24	20	100.0	4	11 US-09-999-686-24	Sequence 24, Appl
25	20	100.0	4	11 US-09-906-393A-1	Sequence 1, Appl
26	20	100.0	4	11 US-09-861-257-19	Sequence 19, Appl
27	20	100.0	4	11 US-09-913-238-63	Sequence 63, Appl
28	20	100.0	4	11 US-09-462-713-13	Sequence 13, Appl
29	20	100.0	4	12 US-09-932-165-1508	Sequence 1508, Ap
30	20	100.0	4	12 US-10-177-725-157	Sequence 157, Ap
31	20	100.0	4	12 US-10-189-360-42	Sequence 42, Appl
32	20	100.0	4	12 US-10-226-877A-33	Sequence 33, Appl
33	20	100.0	4	12 US-10-252-732-8	Sequence 8, Appl
34	20	100.0	4	12 US-10-066-319-1	Sequence 1, Appl
35	20	100.0	4	12 US-10-133-973-63	Sequence 63, Appl
36	20	100.0	4	14 US-10-043-142-9	Sequence 9, Appl
37	20	100.0	4	14 US-10-154-801-3	Sequence 3, Appl
38	20	100.0	4	14 US-10-080-376-77	Sequence 77, Appl
39	20	100.0	4	14 US-10-061-395-25	Sequence 25, Appl
40	20	100.0	4	14 US-10-061-395-35	Sequence 35, Appl
41	20	100.0	4	15 US-10-096-339-24	Sequence 24, Appl
42	20	100.0	4	15 US-10-158-238-17	Sequence 17, Appl
43	20	100.0	4	15 US-10-127-427-2	Sequence 2, Appl
44	20	100.0	4	15 US-10-112-788-18	Sequence 18, Appl
45	20	100.0	4	15 US-10-053-355A-30	Sequence 30, Appl

#### ALIGNMENTS

RESULT 1  
US-08-765-244-16  
; Sequence 16, Application US/08765244  
; Publication No. US20010008771A1  
GENERAL INFORMATION:  
APPLICANT: Seibel, Peter  
TITLE OF INVENTION: CHIMERICAL PEPTIDE-NUCLEIC ACID  
TITLE OF INVENTION: FRAGMENT, PROCESS FOR PRODUCING THE SAME AND ITS USES FOR  
TITLE OR INVENTION: APPROPRIATELY INTRODUCING NUCLEIC ACIDS INTO CELL ORGANELLES  
FILE REFERENCE: 8484-0018-999  
CURRENT APPLICATION NUMBER: US/08/765,244  
PRIOR FILING DATE: 1997-10-30  
PRIOR APPLICATION NUMBER: PCT/DE95/00775  
PRIOR FILING DATE: 1995-06-11  
PRIOR APPLICATION NUMBER: DE P 44 21 079.5  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16  
LENGTH: 4  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic sequence  
US-08-765-244-16

Query Match 100.0%; Score 20; DB 8; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KDEL 4  
|||  
Db 1 KDEL 4

RESULT 2

US-08-873-601-24  
; Sequence 24, Application US/08873601  
; Publication No. US20020064798A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20020064798A1at. Garry P.  
; APPLICANT: Payan, Donald  
; TITLE OF INVENTION: COMBINATORIAL ENZYMAIC COMPLEXES  
; FILE REFERENCE: A-63915/DJB/RMS  
; CURRENT APPLICATION NUMBER: US/08/873,601  
; CURRENT FILING DATE: 1997-06-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: UNKNOWN  
US-08-873-601-24  
Query Match 100.0%; Score 20; DB 8; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KDEL 4  
|||  
Db 1 KDEL 4  
|||  
RESULT 3  
US-09-157-748-30  
; Sequence 30, Application US/09157748  
; Patent No. US20010003042A1  
; GENERAL INFORMATION:  
; APPLICANT: Lorens, James  
; TITLE OF INVENTION: Multiparameter FACS Assays to Detect Alterations in  
; TITLE OF INVENTION: Cell Cycle Regulation  
; FILE REFERENCE: A66587/DJB/RMS  
; CURRENT APPLICATION NUMBER: US/09/157,748  
; CURRENT FILING DATE: 1998-09-21  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 30  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; PUBLICATION INFORMATION:  
; JOURNAL: Royal Society London Transaction B  
; PAGES: 1-1-  
; DATE: 1992  
; US-09-157-748-30  
Query Match 100.0%; Score 20; DB 9; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KDEL 4  
|||  
Db 1 KDEL 4  
|||  
RESULT 4  
US-09-759-960-23  
; Sequence 23, Application US/09759960  
; Patent No. US20010006639A1  
; GENERAL INFORMATION:  
; APPLICANT: Urban, Robert G.  
; APPLICANT: Chicz, Roman M.  
; APPLICANT: Collins, Edward J.  
; APPLICANT: Hedley, Mary Lynn  
; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES FROM THE HPV E7  
; TITLE OF INVENTION: PROTEIN  
; NUMBER OF SEQUENCES: 33

; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson, P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/759,960  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/169,425  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fraezer, Janis K.  
; REGISTRATION NUMBER: 34,819  
; REFERENCE/DOCKET NUMBER: 08191/004002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-542-5070  
; TELEFAX: 617-543-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-759-960-23

Query Match 100.0%; Score 20; DB 9; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KDEL 4  
|||  
Db 1 KDEL 4  
|||  
RESULT 5  
US-09-854-122-2  
; Sequence 2, Application US/09854122  
; Patent No. US20020016980A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTE, RANDALL S.  
; APPLICANT: SMITH, ROBERT  
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PHA-007 01  
; CURRENT APPLICATION NUMBER: US/09/854,122  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/202,529  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide  
US-09-854-122-2  
Query Match 100.0%; Score 20; DB 9; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KDEL 4  
|||

Db 1 KDEL 4

## RESULT 6

US-09-347-064-35  
Sequence 35, Application US/09347064A  
Patent No. US20020045208A1  
GENERAL INFORMATION:  
APPLICANT: Eck, Jurgen  
APPLICANT: Schmidt, Arno  
APPLICANT: Zinke, Holger  
TITLE OF INVENTION: Recombinant Fusion Proteins Based on  
TITLE OF INVENTION: Ribosome-Inactivating Proteins of the mistletoe Viscum  
FILE REFERENCE: 09282-5  
CURRENT APPLICATION NUMBER: US/09/347,064A  
EARLIER FILING DATE: 1999-07-02  
EARLIER APPLICATION NUMBER: PCT/EP98/00009  
EARLIER FILING DATE: 1998-01-02  
EARLIER APPLICATION NUMBER: EP 97 10 0012.0  
EARLIER FILING DATE: 1997-01-02  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 35  
LENGTH: 4  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Modulator  
US-09-347-064-35

Query Match 100.0%; Score 20; DB 9; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.4e+05; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KDEL 4

Db 1 KDEL 4

## RESULT 7

US-09-333-527-2  
Sequence 2, Application US/09333527  
Patent No. US20020078472A1  
GENERAL INFORMATION:  
APPLICANT: Paul CHRISTOU; Eva STROGER; Rainer FISCHER; Carmen MARTIN-VAQUERO; Stef  
TITLE OF INVENTION: METHODS AND MEANS FOR EXPRESSION OF MAMMALIAN POLYPEPTIDES  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski L.L.P.  
STREET: 666 Fifth Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/333,527  
FILING DATE: Concurrently Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/089,322  
FILING DATE: June 15, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary Anne Schofield  
REGISTRATION NUMBER: 36,669  
REFERENCE/DOCKET NUMBER: KL/JIC 202.1 - JEL  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 318-3000

TELEFAX: (212) 752-5958

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 4

TYPE: amino acid

TOPOLOGY: linear

US-09-333-527-2

Query Match 100.0%; Score 20; DB 9; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.4e+05; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KDEL 4

Db 1 KDEL 4

## RESULT 8

US-09-785-921A-16  
Sequence 16, Application US/09785921A  
Patent No. US20020094334A1  
GENERAL INFORMATION:  
APPLICANT: Keener, William K.  
APPLICANT: Ward, Thomas E.  
TITLE OF INVENTION: SELECTIVE DESTRUCTION OF CELLS INFECTED WITH HUMAN  
FILE REFERENCE: LIT-PI-529  
CURRENT APPLICATION NUMBER: US/09/785,921A  
CURRENT FILING DATE: 2001-02-15  
NUMBER OF SEQ ID NOS: 17  
SEQ ID NO 16  
LENGTH: 4  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-785-921A-16

Query Match 100.0%; Score 20; DB 9; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.4e+05; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KDEL 4

Db 1 KDEL 4

## RESULT 9

US-09-997-956-2  
Sequence 2, Application US/09997956  
Patent No. US20020106714A1  
GENERAL INFORMATION:  
APPLICANT: Jalink, Kees  
TITLE OF INVENTION: Membrane Molecule Indicator Compositions  
FILE REFERENCE: P-NS 5045  
CURRENT APPLICATION NUMBER: US/09/997,956  
CURRENT FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/250,679  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/256,559  
PRIOR FILING DATE: 2000-12-18  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PASCSEQ for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 4  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-997-956-2

Query Match 100.0%; Score 20; DB 10; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.4e+05; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 1 KDEL 4

## RESULT 10

US-09-781-804-2  
; Sequence 2, Application US/09781804  
; Patent No. US20020107189A1  
; GENERAL INFORMATION:  
; APPLICANT: ARID Gene Therapeutics, Inc.  
; TITLE OF INVENTION: Regulation of Biological Events Using No. US20020107189A1 Comp  
; FILE REFERENCE: 374 USD1  
; CURRENT APPLICATION NUMBER: US/09/781,804  
; CURRENT FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 64  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: organelle binding domain  
; NAME/KEY: DOMAIN  
; LOCATION: (1)..(4)  
; OTHER INFORMATION: organelle binding domain  
US-09-781-804-2

Query Match 100.0%; Score 20; DB 10; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.4e+05; Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 1 KDEL 4

## RESULT 11

US-09-916-940-28  
; Sequence 28, Application US/09916940  
; Patent No. US20020127564A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20020127564A1a1, Garry P  
; TITLE OF INVENTION: METHODS FOR SCREENING FOR TRANSDOMINANT INTRACELLULAR  
; FILE REFERENCE: A-64260-6/RMS/AMS  
; CURRENT APPLICATION NUMBER: US/09/916,940  
; CURRENT FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: US 09/727,715  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: US 08/963,368  
; PRIOR FILING DATE: 1997-11-03  
; PRIOR APPLICATION NUMBER: US 08/589,109  
; PRIOR FILING DATE: 1996-01-23  
; PRIOR APPLICATION NUMBER: US 08/589,911  
; PRIOR FILING DATE: 1996-01-23  
; PRIOR APPLICATION NUMBER: US 08/789,333  
; PRIOR FILING DATE: 1997-01-23  
; PRIOR APPLICATION NUMBER: US 08/787,738  
; PRIOR FILING DATE: 1997-01-23  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 28  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: endoplasmic  
; OTHER INFORMATION: reticulum sequence.  
US-09-916-940-28

Query Match 100.0%; Score 20; DB 10; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.4e+05;

Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 KDEL 4  
|||  
Db 1 KDEL 4

## RESULT 12

US-09-480-236-8  
; Sequence 8, Application US/09480236  
; Patent No. US20020142000A1  
; GENERAL INFORMATION:  
; APPLICANT: Digan, Mary Ellen  
; APPLICANT: Lake, Philip  
; APPLICANT: Wright, Richard M.  
; TITLE OF INVENTION: Anti-CD3 Immunotoxins and Therapeutic Uses Therefor  
; FILE REFERENCE: CCG 4-31157A/USN  
; CURRENT APPLICATION NUMBER: US/09/480,236  
; CURRENT FILING DATE: 2000-01-10  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PE peptide  
US-09-480-236-8

Query Match 100.0%; Score 20; DB 10; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.4e+05; Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 1 KDEL 4

## RESULT 13

US-09-984-183-3  
; Sequence 3, Application US/09984183  
; Patent No. US20020142983A1  
; GENERAL INFORMATION:  
; APPLICANT: AGRAMAL, BABITA  
; APPLICANT: LONGENECKER, MICHAEL B.  
; TITLE OF INVENTION: MUC-1 ANTAGONISTS AND METHODS OF TREATING IMMUNE  
; FILE REFERENCE: 042881/0130  
; CURRENT APPLICATION NUMBER: US/09/984,183  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 09/457,354  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 60/111,973  
; PRIOR FILING DATE: 1998-12-11  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Retrograde  
; OTHER INFORMATION: transport peptide  
US-09-984-183-3

Query Match 100.0%; Score 20; DB 10; Length 4;

Best Local Similarity 100.0%; Pred. No. 4.4e+05; Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 1 KDEL 4



## RESULT 14

US-09-999-745-22  
; Sequence 22, Application US/09999745  
; Patent No. US20020157120A1  
; GENERAL INFORMATION:  
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: Tsien, Roger Y.  
; APPLICANT: Baird, Geoffrey  
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS  
; FILE REFERENCE: REGEN1470-1  
; CURRENT APPLICATION NUMBER: US/09/999,745  
; PRIOR FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: 09/316,920  
; PRIOR FILING DATE: 1999-05-21  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-999-745-22

Query Match 100.0%; Score 20; DB 10; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 1 KDEL 4

## RESULT 15

US-09-967-772-10  
; Sequence 10, Application US/09967772  
; Patent No. US20020164577A1  
; GENERAL INFORMATION:  
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
; APPLICANT: TSJEN, Roger  
; APPLICANT: GONZALEZ, Jesus  
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS  
; FILE REFERENCE: REGEN1290-5  
; CURRENT APPLICATION NUMBER: US/09/967,772  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: US 09/459,956  
; PRIOR FILING DATE: 1999-12-13  
; PRIOR APPLICATION NUMBER: US 08/765,860  
; PRIOR FILING DATE: 1996-12-19  
; PRIOR APPLICATION NUMBER: PCT/ US96/09652  
; PRIOR FILING DATE: 1996-06-06  
; PRIOR APPLICATION NUMBER: US 08/481,977  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 4  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Protein localization sequence  
US-09-967-772-10

Query Match 100.0%; Score 20; DB 10; Length 4;  
Best Local Similarity 100.0%; Pred. No. 4.4e+05;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 1 KDEL 4

Search completed: August 26, 2003, 16:07:07  
Job time : 2.41593 secs

This page Blank (usps)  
Photo

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 15:54:13 ; Search time 1.80531 Seconds  
(without alignments)  
213.080 Million cell updates/sec

Title: US-09-696-872-37

Perfect score: 20

Sequence: 1 KDEL 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	20	100.0	35	2	JX0200
2	20	100.0	38	2	A38869
3	20	100.0	38	2	C60657
4	20	100.0	38	2	B60657
5	20	100.0	41	2	B37188
6	20	100.0	46	2	S39239
7	20	100.0	50	2	S19624
8	20	100.0	50	2	AH2496
9	20	100.0	52	2	S19623
10	20	100.0	54	2	B96530
11	20	100.0	58	2	A35416
12	20	100.0	58	2	AD3018
13	20	100.0	60	2	AE2387
14	20	100.0	62	2	D69384
15	20	100.0	63	2	S11883
16	20	100.0	65	2	A12513
17	20	100.0	67	2	JU0399
18	20	100.0	68	2	G97737
19	20	100.0	70	2	AC9866
20	20	100.0	71	2	DB1389
21	20	100.0	72	2	EB1126
22	20	100.0	76	2	DB4167
23	20	100.0	77	2	D69010
24	20	100.0	78	2	SE9531
25	20	100.0	79	1	JN0246
26	20	100.0	79	1	B90352
27	20	100.0	79	2	D75326
28	20	100.0	79	2	AG2544
29	20	100.0	80	2	CG9204

30	20	100.0	81	2	S77318	secretory protein
31	20	100.0	82	2	JT0765	rapid lysin III pr
32	20	100.0	82	2	D64329	hypothetical prote
33	20	100.0	83	2	B82881	hypothetical prote
34	20	100.0	83	2	B82887	hypothetical prote
35	20	100.0	85	2	I40586	hypothetical prote
36	20	100.0	85	2	T39536	probable involveme
37	20	100.0	86	1	WMECSR	replication contro
38	20	100.0	89	2	S05557	hypothetical prote
39	20	100.0	90	2	T04077	probable ribosomal
40	20	100.0	90	2	AC1030	hypothetical prote
41	20	100.0	91	2	E59100	hypothetical prote
42	20	100.0	92	2	H84264	hypothetical prote
43	20	100.0	92	2	H84296	hypothetical prote
44	20	100.0	93	2	T08319	conserved hypothet
45	20	100.0	93	2	AG2250	hypothetical prote

## ALIGNMENTS

## RESULT 1

JX0200  
gurmardin - Gymnema sylvestre  
N:Alternate names: sweet-taste-suppressing peptide  
C:Species: Gymnema sylvestre  
C>Date: 17-Aug-1992 #sequence\_revision 17-Aug-1992 #text\_change 25-Oct-1996  
C:Accession: JX0200  
R:Kamei, K.; Takano, R.; Miyasaka, A.; Imoto, T.; Hara, S.  
J. Biochem. 111, 109-112, 1992  
A>Title: Amino acid sequence of sweet-taste-suppressing peptide (gurmardin) from the leave  
A:Reference number: JX0200; MUID:92299642; PMID:1607357  
A:Accession: JX0200  
A:Molecule type: protein  
A:Residues: 1-35 <KAM>  
A:Experimental source: leaf  
A:Comment: This protein suppresses sweet taste.  
C:Keywords: pyroglutamic acid; sweet taste  
F.I/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 100.0%; Score 20; DB 2; Length 35;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
DB 6 KDEL 9

## RESULT 2

A38869  
circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate 427-5) (Frac  
C:Species: Plasmodium falciparum  
C>Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 09-Jun-2000  
C:Accession: A38869  
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.  
Mol. Biochem. Parasitol. 37, 275-280, 1989  
A>Title: Wild isolates of Plasmodium show extensive polymorphism in T cell epi  
A:Reference number: A60657; MUID:9011334; PMID:2481827  
A:Accession: A38869  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-38 <LOC>  
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match 100.0%; Score 20; DB 2; Length 38;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
DB 28 KDEL 31

```
RESULT 3
C60657
C:circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate 7G8) (fragm
C:Species: Plasmodium falciparum
C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 09-Jun-2000
C:Accession: C60657
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.
Mol. Biochem. Parasitol. 37, 275-280, 1989
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell ep
A:Reference number: A60657; MUID:90114334; PMID:2481827
A:Accession: C60657
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-38 <LOC>
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match          100.0%; Score 20; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 28 KDEL 31

RESULT 4
B60657
C:circumsporozoite protein - malaria parasite (Plasmodium falciparum) (isolate 79-98) (fra
C:Species: Plasmodium falciparum
C:Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 09-Jun-2000
C:Accession: B60657
R:Lockyer, M.J.; Marsh, K.; Newbold, C.I.
Mol. Biochem. Parasitol. 37, 275-280, 1989
A:Title: Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell ep
A:Reference number: A60657; MUID:90114334; PMID:2481827
A:Accession: B60657
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-38 <LOC>
C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

Query Match          100.0%; Score 20; DB 2; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 28 KDEL 31

RESULT 5
B37188
major body fluid allergen ABA-1 - common roundworm (fragment)
C:Species: Ascaris lumbricoles (common roundworm)
C:Date: 31-Jan-1992 #sequence_revision 31-Jan-1992 #text_change 30-Sep-1993
C:Accession: B37188
R:Christie, J.F.; Dunbar, B.; Davidson, I.; Kennedy, M.W.
Immunology 69, 596-602, 1990
A:Title: N-terminal amino acid sequence identity between a major allergen of Ascaris lum
A:Reference number: A37188; MUID:90243308; PMID:2335378
A:Accession: B37188
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-41 <CHR>

Query Match          100.0%; Score 20; DB 2; Length 41;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 21 KDEL 24
```

```
RESULT 6
S39239
C:Shiga enterotoxin chain A2 - Vibrio cholerae (serotype 0139)
N:Alternate names: CTx2 cholera toxin chain A2
C:Species: Vibrio cholerae
A:Variety: serotype 0139
C:Date: 08-May-1995 #sequence_revision 26-Jul-1996 #text_change 16-Jul-1999
C:Accession: S39239; S39240
R:Jedens, M.; Holmgren, U.
Submitted to the EMBL Data Library, November 1993
A:Description: Structure and arrangement of the Cholera toxin genes in vibrio Cholera 01:
A:Reference number: S39238
A:Accession: S39239
A:Molecule type: DNA
A:Residues: 1-46 <LEB>
A:Cross-references: EMBL:X76390; NID:q433856; PIDN:CAAS3974.1; PID:q433858; EMBL:X76391;
A:Experimental source: strain 4260B, serotype 0139
C:Superfamily: heat-labile enterotoxin chain A

Query Match          100.0%; Score 20; DB 2; Length 46;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 43 KDEL 46

RESULT 7
S19624
ornatin B - leech (Placobdella ornata)
C:Species: Placobdella ornata
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Jul-1997
C:Accession: S19624
R:Mazur, P.; Henzel, W.J.; Seymour, J.L.; Lazarus, R.A.
Eur. J. Biochem. 202, 1073-1082, 1991
A:Title: Ornatin: potent glycoprotein IIb-IIIa antagonists and platelet aggregation inh
A:Reference number: S19566; MUID:92111479; PMID:1765068
A:Accession: S19624
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-50 <MAZ>

Query Match          100.0%; Score 20; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
Db 7 KDEL 10

RESULT 8
AH2496
transposase asr7152 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH2496
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriunchi,
Nakazaki, N.; Shimp, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anat
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2496
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-50 <KUR>
A:Cross-references: GB:BA000020; PIDN:BA878236.1; PID:917135690; GSPDB:GN00180
A:Experimental source: strain PCC 7120
C:Genetics:
```

A:Gene: asr7152  
A:Genome: plasmid

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 50;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 8 KDEL 11

RESULT 9  
S19623  
ornatin C - leech (Placobdella ornata)  
C:Species: Placobdella ornata  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 24-Jul-1997  
C:Accession: S19623  
R:Maizur, P.; Henzel, W.J.; Seymour, J.L.; Lazarus, R.A.  
B: J. Biochem. 202, 1073-1082, 1991  
A:Title: Ornatin: potent glycoprotein IIb-IIIa antagonists and platelet aggregation inhibitors  
A:Reference number: S19566; MUID:92111479; PMID:1765068  
A:Accession: S19623  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-52 <MAZ>

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 52;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 7 KDEL 10

RESULT 10  
E96530  
hypochemical protein FljF21.16 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E96530  
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A6141; MUID:21016719; PMID:11130712  
A:Accession: E96530  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-54 <STO>  
A:Cross-references: GB:AB005173; NID:95430759; PIDN:AA43159.1; GSPDB:GN00141  
A:Gene: FljF21.16  
A:Map position: 1

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 54;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 47 KDEL 50

RESULT 11  
A35416

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - bovine (fragments)  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 18-Jan-1991 #sequence\_revision 18-Jan-1991 #text\_change 23-Jun-1993  
C:Accession: A35416  
R:Aucutt, C.A.; Barot, H.A.; Margerison, E.E.C.; Turcatti, G.; Winfield, P.; Hayes, M.V.  
Biochem. Biophys. Res. Commun. 170, 763-768, 1990  
A:Title: Structure and partial amino acid sequence of calf thymus DNA topoisomerase II:  
A:Reference number: A35416; MUID:90343796; PMID:1696476  
A:Accession: A35416  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-58 <AUS>  
C:Keywords: isomerase

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 58;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 55 KDEL 58

RESULT 12  
AD3018  
conserved hypochemical protein Atu3752 [imported] - Agrobacterium tumefaciens (strain C58)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AD3018  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; Erang, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krepsan, W.; Perry, M.; Gordon-Kamm, E. ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AD3018  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-58 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AA144562.1; PID:917742178; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu3752  
A:Map position: linear chromosome

Query Match  
Best Local Similarity 100.0%; Score 20; DB 2; Length 58;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
|||  
Db 46 KDEL 49

RESULT 13  
AE2387  
hypochemical protein asr4653 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AE2387  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AE2387  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-60 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA876352.1; PID:G17133790; GSPDB:GN00179

A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: asr4653

Query Match 100.0%; Score 20; DB 2; Length 60;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
|||  
Db 52 KDEL 55

## RESULT 14

D69384  
hypothetical protein AF1076 - Archaeoglobus fulgidus  
C:Species: Archaeoglobus fulgidus  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: D69384  
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E.F.  
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
Nature 390, 364-370, 1997  
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artiaeh, P.; Kaime, B.P.; Sykes, S.  
Smith, H.O.; Woese, C.R.; Venter, J.C.  
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae  
A:Reference number: A69250; MUID:98049343; PMID:9389475  
A:Accession: D69384  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-62 <KDE>  
A:Cross-references: GB:AE001028; GB:AE000782; NID:g2689351; PIDN:AAB90163.1; PID:g264951

Query Match 100.0%; Score 20; DB 2; Length 62;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
|||  
Db 8 KDEL 11

## RESULT 15

S11883  
copB protein - Escherichia coli plasmid ColV2-K94  
C:Species: Escherichia coli  
C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 30-May-1997  
C:Accession: S11883  
R:Banerjee, A.; Weber, P.C.; Palchaudhuri, S.  
Mol. Gen. Genet. 220, 320-324, 1990  
A:Title: Comparison of the CopB systems of plasmids R1 and ColV2-K94: a single base alte  
A:Reference number: S11883; MUID:90220506; PMID:2325626  
A:Accession: S11883  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-63 <MOL>  
C:Genetics:  
A:Gene: copB  
A:Genome: plasmid ColV2-K94  
C:Superfamily: repA2 protein

Query Match 100.0%; Score 20; DB 2; Length 63;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4  
|||  
Db 53 KDEL 56

Search completed: August 26, 2003, 16:05:26  
Job time : 6.80531 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 26, 2003, 16:01:13 ; Search time 0.9915 Seconds  
(without alignments)  
189.786 Million cell updates/sec

Title: US-09-696-872-37  
Perfect score: 20  
Sequence: 1 KDEL 4

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	28	1	ORND_PLAOR
2	20	100.0	35	1	GUR_GYMSY
3	20	100.0	37	1	VPV_HV128
4	20	100.0	50	1	ORNE_PLAOR
5	20	100.0	52	1	ORNC_PLAOR
6	20	100.0	62	1	YA76_ARCFU
7	20	100.0	67	1	Y00A_BPT4
8	20	100.0	76	1	Y066_HALN1
9	20	100.0	78	1	HOLI_BPH1
10	20	100.0	78	1	SI0D_HUMAN
11	20	100.0	81	1	SECE_SYNY3
12	20	100.0	82	1	VR3_BPT4
13	20	100.0	82	1	Y235_METUA
14	20	100.0	86	1	CPB3_ECOLI
15	20	100.0	89	1	DSBA_BPT4
16	20	100.0	97	1	Y04E_BPT4
17	20	100.0	101	1	ALP_LYMET
18	20	100.0	103	1	YC05_METUA
19	20	100.0	104	1	RECA_LEUME
20	20	100.0	105	1	Y349_METUA
21	20	100.0	105	1	YE95_MYCTU
22	20	100.0	109	1	MYCOCBACTE
23	20	100.0	111	1	RLAI_ARTSA
24	20	100.0	111	1	Y4E8_ENTRA
25	20	100.0	112	1	ANFG_AZONA
26	20	100.0	112	1	RRP3_SYNY3
27	20	100.0	114	1	Y474_METUA
28	20	100.0	114	1	THH1_ARATH
29	20	100.0	116	1	ANFG_CLOPA
30	20	100.0	116	1	ARSC_HAEIN
31	20	100.0	116	1	THIH_FAGES
32	20	100.0	118	1	THH2_TOBAC
33	20	100.0	118	1	THIH_RICCO

34	20	100.0	118	1	V118_ASFB7	P18556 african swi
35	20	100.0	120	1	RL31_PICMA	O65071 picea maria
36	20	100.0	121	1	PPDB_METTH	O26774 methanobact
37	20	100.0	123	1	RL7A_METKA	O8tv03 methanopyru
38	20	100.0	123	1	RL7A_METTH	O26355 methanobact
39	20	100.0	123	1	RL7A_YEAST	P19955 saccharomyc
40	20	100.0	126	1	RK12_CYAPA	P48124 cyanophora
41	20	100.0	127	1	UCR1_YEAST	P00128 saccharomyc
42	20	100.0	128	1	RL7_CORGL	O8nt28 corynebacte
43	20	100.0	129	1	RS8_ECOLI	P02361 escherichia
44	20	100.0	129	1	RS8_HAEIN	P44377 haemophilus
45	20	100.0	130	1	RS8_BUCAK	P46180 buchnera ap

## ALIGNMENTS

RESULT 1	ID	ORND_PLAOR	STANDARD	PRT	28 AA.
AC	P25513				
DT	01-MAY-1992 (Rel. 22, Created)				
DT	01-MAY-1992 (Rel. 22, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Ornatin D (Fragment).				
OS	Placodella ornata (Turtle leech).				
OC	Eukaryota; Metazoa; Annelida; Clitellata; Hirudindida; Hirudinea;				
OC	Rhynchobdellida; Glossiphoniidae; Placodella.				
OX	NCBI_TaxID=6415;				
RN	[1]				
RP	SEQUENCE.				
RX	MEDLINE=92111479; PubMed=1765068;				
RA	Mazur P., Henzel W.J., Seymour J.L., Lazarus R.A.;				
RT	"Ornatins: potent glycoprotein IIB-IIIA antagonists and platelet				
RT	aggregation inhibitors from the leech Placodella ornata.";				
RL	Eur. J. Biochem. 202:1073-1082 (1991).				
CC	-I- FUNCTION: POTENT INHIBITOR OF FIBRINOGEN INTERACTION WITH PLATELET				
CC	RECEPTORS EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. MAY PREVENT				
CC	BLOOD FROM CLOTTING DURING EITHER FEEDING AND/OR STORAGE OF				
CC	INGESTED BLOOD.				
CC	-I- SIMILARITY: BELONGS TO THE ORNATIN FAMILY.				
DR	InterPro: IPR002463; Ornatin.				
KW	Pfam: PF02088; Ornatin; 1.				
KW	Blood coagulation; Platelet; Cell adhesion.				
FT	NON TER				
FT	SEQUENCE 28 AA; 3361 MW; CFC38951F9137C2 CRC64;				
Qy	Query Match				
Qy	Best Local Similarity 100.0%; Score 20; DB 1; Length 28;				
Qy	Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1 KDEL 4				
Qy	7 KDEL 10				
Db					
RESULT 2					
ID	GUR_GYMSY	STANDARD;	PRT;	35 AA.	
AC	P25810;				
DT	01-MAY-1992 (Rel. 22, Created)				
DT	01-FEB-1984 (Rel. 28, Last sequence update)				
DT	15-SEP-2003 (Rel. 42, Last annotation update)				
DE	Gurmarin (Sweet-taste-suppressing peptide).				
OS	Gymnema sylvestre (Gurmar).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
OC	Asteridae; Lamiales; Gentianales; Apocynaceae; Periplocoideae;				
OC	Gymnema.				
OX	NCBI_TaxID=4068;				
OX	[1]				
RP	SEQUENCE.				
RP	TISSUE=leaf;				

RX MEDLINE=92299642; PubMed=1607357;  
 RA Kami K., Takano R., Miyasaka A., Imoto T., Hara S.;  
 RT "Amino acid sequence of sweet-taste-suppressing peptide (gummarin)  
 from the leaves of *Gymnema sylvestre*.";  
 RL J. Biochem. 111:109-112(1992).  
 RN [2]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=95307084; PubMed=7787425;  
 RA Arai K., Ichima R., Morikawa S., Miyasaka A., Imoto T., Yoshimura S.,  
 RA Aimoto S., Akasaka K.;  
 RT "Three-dimensional structure of gummarin, a sweet taste-suppressing  
 polypeptide.";  
 RL J. Biomol. NMR 5:297-305(1995).  
 RN [3]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=99421659; PubMed=10491100;  
 RA Fletcher J.I., Dingley A.J., Smith R., Connor M., Christie M.J.,  
 RA King G.F.;  
 RT "High-resolution solution structure of gummarin, a sweet-taste-  
 suppressing plant polypeptide.";  
 RL Eur. J. Biochem. 264:525-533(1999).  
 CC -!- FUNCTION: SUPPRESSES STRONGLY THE SWEET TASTE RESPONSES IN THE RAT  
 WITH HIGH SPECIFICITY TO SUCROSE, GLUCOSE, GLYCINE, AND SACCCHARIN.  
 THIS EFFECT IS REVERSIBLE, BUT COMPLETE RECOVERY OF THE SUPPRESSED  
 RESPONSES REQUIRED AT LEAST 3H. GUMMARIN SHOWED NO EFFECT OR ONLY  
 A VERY WEAK EFFECT ON THE SWEET TASTE SENSATION IN HUMANS.  
 CC PIR: JX0200; JX0200.  
 DR PDB: 1GUR; 01-AUG-96.  
 DR PDB: 1GAE; 27-AUG-96.  
 KW Taste-modifying protein; 3D-structure; Pyrrolidone carboxylic acid.  
 KM MOD\_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.  
 FT DISULFID 3 18  
 FT DISULFID 10 23  
 FT DISULFID 17 33  
 FT TURN 12 15  
 FT STRAND 24 27  
 FT TURN 28 29  
 FT STRAND 30 32  
 SQ SEQUENCE 35 AA; 4232 MW; 661BD6FBAF504CB6 CRC64;  
 QY Query Match 100.0%; Score 20; DB 1; Length 35;  
 Best Local Similarity 100.0%; Pred. No. 81;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 Db 6 KDEL 9  
 RESULT 3  
 VPU\_HV128  
 ID VPU\_HV128 STANDARD; PRT; 37 AA.  
 AC P08807;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE VPU protein (U ORF protein) (Fragment).  
 GN VPU  
 OS Human immunodeficiency virus type 1 (Z-84 isolate) (HIV-1).  
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
 NCBI\_TaxID=11681;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88281278; PubMed=3395517;  
 RA Yourno J., Joseph S.F., Reitz M.S. Jr., Zagury D., Wong-Staal F.,  
 RA Gallo R.C.;  
 RT "Nucleotide sequence analysis of the env gene of a new Zairian  
 isolate of HIV-1.";  
 RL Aids Res. Hum. Retroviruses 4:165-173(1988).  
 CC -!- FUNCTION: ACTS IN THE DEGRADATION OF CD4 IN THE ENDOPLASMIC  
 RETICULUM AND IN THE ENHANCEMENT OF VIRION RELEASE FROM THE PLASMA  
 MEMBRANE OF INFECTED CELLS.  
 CC -!- SUBCELLULAR LOCATION: Membrane-bound.

CC -!- MISCELLANEOUS: THE Z-84 ISOLATE WAS TAKEN FROM A 54 YEAR-OLD  
 CC ZAIREAN MALE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib.ch](mailto:license@sib.ch)).  
 CC -----  
 CC EMBL: J03653; AAA44683.1; -.  
 DR HSSP: P19554; IVPV.  
 DR HIV: J03653; VPU5JY1.  
 DR InterPro: IPR002094; Vpu.  
 DR Pfam: PF00558; Vpu; 1.  
 KW Transmembrane; AIDS.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 37 AA; 4131 MW; 00EA53CA27D09509 CRC64;  
 QY Query Match 100.0%; Score 20; DB 1; Length 37;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 Db 17 KDEL 20  
 RESULT 4  
 ORNE\_PLAOR  
 ID ORNE\_PLAOR STANDARD; PRT; 50 AA.  
 AC P2514;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ornatin E.  
 DE Ornatin E.  
 OS Placobdella ornata (Turtle leech).  
 OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudindae; Hirudinea;  
 OC Rhynchobdellida; Glossiphoniidae; Placobdella.  
 NCBI\_TaxID=6415;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=9211479; PubMed=1765068;  
 RA Mazur P., Henzel W.J., Seymour J.L., Lazarus R.A.;  
 RT "Ornatin: potent glycoprotein IIB-IIIA antagonists and platelet  
 aggregation inhibitors from the leech *Placobdella ornata*.";  
 RL Eur. J. Biochem. 202:1073-1082(1991).  
 CC -!- FUNCTION: POTENT INHIBITOR OF FIBRINOGEN INTERACTION WITH PLATELET  
 CC RECEPTORS EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. MAY PREVENT  
 CC BLOOD FROM CLOTTING DURING EITHER FEEDING AND/OR STORAGE OF  
 CC INGESTED BLOOD.  
 CC -!- SIMILARITY: BELONGS TO THE ORNATIN FAMILY.  
 DR PIR: S19624; S19624.  
 DR InterPro: IPR002463; Ornatin.  
 DR Pfam: PF02088; Ornatin; 1.  
 DR Prodom: PD012062; Ornatin; 1.  
 KW Blood coagulation; Platelet; Cell adhesion.  
 FT SITE 42 44 CELL ATTACHMENT SITE.  
 SQ SEQUENCE 50 AA; 5727 MW; 688408EB1E00192 CRC64;  
 QY Query Match 100.0%; Score 20; DB 1; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 12e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 Db 7 KDEL 10  
 RESULT 5  
 ORNC\_PLAOR  
 ID ORNC\_PLAOR STANDARD; PRT; 52 AA.



```

AC P25512;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ornatin C.
OS Placobdella ornata (Turtle leech).
OC Eukaryota; Metazoa; Annelida; Clitellata; Hirudindae; Hirudinea;
OC Rhynchobdellida; Glossiphoniidae; Placobdella.
OX NCBI_TaxID=6415;
RN [1]
RP SEQUENCE.
RX MEDLINE=92111479; PubMed=1765068;
RA Maun P., Henzel W.J., Seymour J.L., Lazarus R.A.;
RT "Ornatins: potent glycoprotein iib-iiia antagonists and platelet
RT aggregation inhibitors from the leech Placobdella ornata.";
RL Eur. J. Biochem. 202:1073-1082(1991)
CC -1- FUNCTION: POTENT INHIBITOR OF FIBRINOGEN INTERACTION WITH PLATELET
CC RECEPTORS EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. MAY PREVENT
CC BLOOD FROM CLOTTING DURING EITHER FEEDING AND/OR STORAGE OF
CC INGESTED BLOOD.
CC -1- SIMILARITY: BELONGS TO THE ORNATIN FAMILY.
DR PIR: S19623; S19623.
DR InterPro; IPR002463; Ornatin.
DR Pfam; PF02088; Ornatin; 1.
DR ProDom; PD012062; Ornatin; 1.
KW Blood coagulation; Platelet; Cell adhesion.
FT SITE 42 44
FT SEQUENCE 52 AA; 5845 MW; BA55CA7408E4F09 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
DB 7 KDEL 10

RESULT 6
YA76_ARCFU STANDARD; PRT; 62 AA.
ID YA76_ARCFU
AC O29187;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF1076.
GN AF1076.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2224;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
KLANK H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrleides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kitzner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Spriggs T., Weidman J.F., McDonald L., Uterback T.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Maeson T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001028; AAB90163.1; -.
DR PIR; D69384; D69384.
DR TIGR; AF1076; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 62 AA; 7385 MW; E2F541BA57A10477 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 62;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
DB 8 KDEL 11

RESULT 7
Y00A_BPT4 STANDARD; PRT; 67 AA.
ID Y00A_BPT4
AC P25186;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 8.1 kDa protein in rIIA-gp60 intergenic region.
GN Y00A OR RIIA.1.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like viruses.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90337270; PubMed=2379817;
RA Daegelen P., Brody E.;
RT "The rIIA gene of bacteriophage T4. I. Its DNA sequence and discovery
RT of a new open reading frame between genes 60 and rIIA.";
RL Genetics 125:237-248(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Kuter E., Arisaka F., Kunitawa T., Tsugita A., Mosig G.,
RA Mezyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52686; CA36910.1; -.
DR EMBL; AF158101; AAD42648.1; -.
DR PIR; J00399; J00399.
KW Hypothetical protein.
SQ SEQUENCE 67 AA; 8129 MW; B2DAA25C721A8753 CRC64;

Query Match 100.0%; Score 20; DB 1; Length 67;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4
DB 52 KDEL 55

RESULT 8
Y066_HALN1 STANDARD; PRT; 76 AA.
ID Y066_HALN1

```

```

AC Q9HSU8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein Vng0066h.
GN VNG0066h.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxId=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Laskey S.R., Baliga N.S., Thorson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welt R., Goo Y.A.,
RA Lettner D.G., Jablonki P.B., Krebs M.P., Angvine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlischroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -1- SIMILARITY: BELONGS TO THE UPF0175 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB004976; AAG18704.1; -.
DR PIR; D84167; D84167.
DR InterPro; IPR005368; UPF0175.
DR Pfam; PF03683; UPF0175; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 76 AA; 8577 MW; 4083B6DCFA6C87B CRC64;

Query Match          100.0%; Score 20; DB 1; Length 76;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
   ||||
Db 47 KDEL 50

RESULT 9
HOLI_BPHPI
ID HOLI_BPHPI STANDARD; PRT; 78 AA.
AC P51727;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HOLI.
GN HOLI.
OS Bacteriophage HPI.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.
OX NCBI_TaxId=10690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85128433; PubMed=6098523;
RA Benjamin R.C., Fitzmaurice W.P., Huang P.C., Socca J.J.;
RT "Nucleotide sequence of cloned DNA segments of the Haemophilus
RT influenzae bacteriophage HPIc1."
RL Gene 31:173-185(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=HPIc1;
RL MEDLINE=96279738; PubMed=8710508;

```

```

RA Eposito D., Fitzmaurice W.P., Benjamin R.C., Goodman S.D.,
RA Waldman A.S., Socca J.J.;
RT "The complete nucleotide sequence of bacteriophage HPI DNA."
RL Nucleic Acids Res. 24:2360-2368(1996).
CC -1- FUNCTION: INDUCES HOST CELL LYSIS BY CAUSING FORMATION OF PORES IN
CC THE INNER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U24159; AB09210.1; -.
DR PIR; S69531; S69531.
KW Phage lysin protein; Transmembrane; Inner membrane.
FT TRANSMEM 8 24
SQ SEQUENCE 78 AA; 8694 MW; 3FC0596BEOA4FED CRC64;

Query Match          100.0%; Score 20; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
   ||||
Db 55 KDEL 58

RESULT 10
SID_HUMAN
ID SID_HUMAN STANDARD; PRT; 78 AA.
AC P29377;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vitamin D-dependent calcium-binding protein, intestinal (CABP)
DE (Calbindin D9k).
GN CALB3 OR S100D OR CABP9K.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92304291; PubMed=1610358;
RA Howard A., Legon S., Spurr N.K., Walters J.R.I.;
RT "Molecular cloning and chromosomal assignment of human calbindin-
RT D9k."
RL Biochem. Biophys. Res. Commun. 185:663-669(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92354716; PubMed=1379540;
RA Jeung B.B., Krisinger J., Dann J.L., Jeung P.C.K.;
RT "Molecular cloning of the full-length cDNA encoding the human
RT calbindin-D9k."
RL FEBS Lett. 307:224-228(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94141916; PubMed=8308886;
RA Jeung B.B., Jeung P.C.K., Krisinger J.;
RT "The human calbindin-D9k gene: Complete structure and implications on
RT steroid hormone regulation."
RL J. Mol. Biol. 235:1231-1238(1994).
CC -1- SIMILARITY: BELONGS TO THE S-100 FAMILY.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL, X65869; CAA46699.1; -  
CC EMBL, L13420; AAA35638.1; -  
CC EMBL, L13042; AAA35637.1; -  
CC PIR, JN0246; JN0246.  
CC HSSP; P02632; ICBI.  
CC Genew; HGNC:1436; CALB3.  
CC MIM; 302020; -  
CC GO; GO:0005509; F:calcium ion binding activity; TAS.  
CC InterPro; IPR001751; CAPB\_S100.  
CC InterPro; IPR002048; EF-hand.  
CC Pfam; PF00035; ehand; 1.  
CC Pfam; PF01023; S\_100; 1.  
CC ProDom; PD003407; CAPB\_S100; 1.  
CC ProDom; PD000012; EF-hand; 1.  
CC PROSITE; PS00018; EF-HAND; 1.  
CC PROSITE; PS00303; S100 CAPB; 1.  
CC Calcium-binding; Vitamin D; Acetylation.  
CC INIT\_MBT 0 0 BY SIMILARITY.  
CC MOD\_RES 1 1 ACETYLATION (BY SIMILARITY).  
CC CA\_BIND 17 30 EF-HAND 1 (LOW AFFINITY).  
CC CA\_BIND 57 68 EF-HAND 2 (HIGH AFFINITY).  
CC CONFLICT 78 78 O -> S (IN REF. 3).  
CC SEQUENCE 78 AA; 8885 MW; 110FA5B92485B9F6 CRC64;  
-----  
Query Match 100.0%; Score 20; DB 1; Length 78;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
OY 1 KDEL 4  
DB 28 KDEL 31  
-----  
RESULT 11  
SECE\_SYNY3  
ID SECE\_SYNY3 STANDARD; PRT; 81 AA.  
AC P38382;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Preprotein translocase secE subunit.  
GN SECE OR SCL335.  
OS Synechocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
OX NCBI\_TaxID=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93281410; PubMed=7685084;  
RA Schmidt J., Subramanian A.R.;  
RT "Sequence of the cyanobacterial rRNA(w) gene in Synechocystis PCC  
RT 6803: requirement of enzymatic 3' CCA attachment to the acceptor-  
RT stem";  
RL Nucleic Acids Res. 21:2519-2519(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneo T., Sato S., Koreani H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K.,  
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,  
RA Yamada M., Yasuda M., Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions";  
RL DNA Res. 3:109-116(1996).  
CC -1- FUNCTION: ESSENTIAL FOR PROTEIN EXPORT.  
CC -1- SUBCELLULAR LOCATION: Tail-anchored membrane protein (Potential).  
CC -1- SIMILARITY: Belongs to the sece/SEC61-gamma family.

CC -----  
CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL, X72627; -; NOT ANNOTATED\_CDS.  
CC EMBL, D90906; BAA17421.1; -  
CC PIR; S77318; S77318.  
CC InterPro; IPR001901; SECE.  
CC InterPro; IPR005807; SECE\_bac.  
CC Pfam; PF00584; SECE; 1.  
CC TIGRFAMs; TIGR00964; 3a0501806; 1.  
CC PROSITE; PS01067; SECE\_SEC61G; 1.  
CC Protein transport; Translocation; Transmembrane; Complete proteome.  
CC TRANSMEM 50 70 POTENTIAL.  
CC SEQUENCE 81 AA; 9134 MW; C3D3E08A8100209F CRC64;  
-----  
Query Match 100.0%; Score 20; DB 1; Length 81;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
-----  
OY 1 KDEL 4  
DB 31 KDEL 34  
-----  
RESULT 12  
VR3\_BPT4  
ID VR3\_BPT4 STANDARD; PRT; 82 AA.  
AC P17309;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Protein rIII.  
GN RIII OR 31..1.  
OS Bacteriophage T4.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
OC T4-like viruses.  
OX NCBI\_TaxID=10665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D;  
RX MEDLINE=90301484; PubMed=2362813;  
RA Prilipov A.G., Mesyanzhinov V.V., Aebi U., Kellenberger E.;  
RT "Cloning and sequencing of bacteriophage T4 genes between map  
RT positions 128.3-130.3";  
RL Nucleic Acids Res. 18:3635-3635(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90332452; PubMed=2377483;  
RA Raudoniklene A., Nivinskas R.;  
RT "Nucleotide sequence of bacteriophage T4 gene 31 region";  
RL Nucleic Acids Res. 18:4280-4280(1990).  
RN [3]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RX MEDLINE=92267389; PubMed=1587487;  
RA Raudoniklene A., Nivinskas R.;  
RT "Gene rIII is the nearest downstream neighbour of bacteriophage T4  
RT gene 31.";  
RL Gene 114:85-90(1992).  
RN [4]  
RP SEQUENCE FROM N.A., AND MUTANTS.  
RX MEDLINE=94063508; PubMed=8244025;  
RA Raudoniklene A., Nivinskas R.;  
RT "The sequences of gene rIII of bacteriophage T4 and its mutants";  
RL Gene 134:135-136(1993).  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Kutter E., Aizsaka F., Kunisawa T., Tsugita A., Mosig G.,

RA Mesyanzhinov V., Ruger W., Stihham T., Thomas E.;  
 RL "Bacteriophage T4 genome analysis";  
 RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X17657; CAA35652.1; -  
 DR EMBL; M37882; AAA32507.1; -  
 DR EMBL; X54536; CAA38406.1; -  
 DR EMBL; AF158101; AAD42650.1; -  
 DR PIR; J70765; J70765. -  
 FT VARIANT 42 42  
 FT  
 FT VARIANT 82 82 H -> R (IN R67 MUTANT; R-TYPE PLAQUE  
 FT MORPHOLOGY).  
 FT K -> E (IN RES40 AND RBB9 MUTANTS; R-TYPE  
 FT PLAQUE MORPHOLOGY).  
 FT CONFLICT 64 82 KPLASARKAVRHFFVTLK -> NMLLHKEQFTLW  
 FT (IN REF. 1).  
 FT  
 FT SEQUENCE 82 AA; 9325 MW; CFE8DF45AB9CC2CD CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 ID |||||  
 DB 57 KDEL 60  
 ID Y235\_METUA STANDARD; PRT; 82 AA.  
 Y235\_METUA  
 AC Q57687;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ0235.  
 OS *Mycobacterium tuberculosis*  
 GN MJ0235.  
 NC Metanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_Taxid=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weisscock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geisheiden N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
 RA Utermbeck T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Frazer C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, *Methanococcus*  
 RA *jannaschii*.";  
 RT Science 273:1058-1073(1996).  
 RL -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U67479; AAB88227.1; -  
 DR PIR; D64329; D64329. -

DR TIGR; MJ0235; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 82 AA; 9899 MW; CB23BFED0A7F4B9F CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 ID |||||  
 DB 36 KDEL 39  
 ID CPB3\_ECOLI STANDARD; PRT; 86 AA.  
 CPB3\_ECOLI  
 AC P03855;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Replication regulatory protein repA2 (copB protein).  
 GN REP2A OR COPB OR REPB OR CP0258.  
 OS *Escherichia coli*, and  
 OS *Shigella flexneri*.  
 OG Shigella flexneri.  
 OG Plasmid IncFII R1-19 (R1 dtd-19), and Plasmid pCP301.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_Taxid=562, 623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=E.coli; PLASMID=IncFII R1-19 (R1 dtd-19);  
 RX MEDLINE=81172236; PubMed=6261081;  
 RA Stougaard P., Molin S., Nordstrom K., Hansen F.G.;  
 RT "The nucleotide sequence of the replication control region of the  
 RT resistance plasmid R1dtd-19.";  
 RL Mol. Gen. Genet. 181:116-122(1981).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a; PLASMID=pCP301;  
 RX MEDLINE=22272406; PubMed=12384590;  
 RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
 RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
 RA Sun L., Xue Y., Zhao A., Gao Y., Kan B., Ding K., Chen S.,  
 RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
 RA Yu J.;  
 RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
 RT through comparison with genomes of *Escherichia coli* K12 and O157.";  
 RL Nucleic Acids Res. 30:4432-4441(2002).  
 CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE DETERMINATION OF COPY  
 CC NUMBER IN GENE REPLICATION. IT BINDS TO THE REPA PROMOTER THUS  
 CC INHIBITING THE SYNTHESIS OF THE MRNA FOR THE INITIATOR PROTEIN  
 CC REPA.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF386526; AAL72557.1; -  
 DR PIR; A04484; WMECSR.  
 KW Plasmid; Plasmid copy control; Transcription regulation; Repressor;  
 KW DNA-binding.  
 SQ SEQUENCE 86 AA; 10038 MW; F7AD85B2C211CBF8 CRC64;  
 Query Match 100.0%; Score 20; DB 1; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDEL 4  
 ID |||||

Db 76 KDEL 79

## RESULT 15

DSBA\_BPT4 STANDARD; PRT; 89 AA.  
 ID\_DSBA\_BPT4  
 AC P13320;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Double-stranded binding protein (dsDNA binding protein A).  
 GN DSBA OR RPB8.  
 OS Bacteriophage T4.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC T4-like viruses.  
 OX NCBI\_TaxId=10665;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BK536;  
 RX MEDLINE=89386003; PubMed=2674900;  
 RA Hahn S., Rueger W.;  
 RT "Organization of the bacteriophage T4 genome between map positions  
 150,745 and 145,824.";  
 RL Nucleic Acids Res. 17:6729-6729 (1989).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Kutter E., Arisaka F., Kunitawa T., Tsugita A., Moesig G.,  
 RA Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;  
 RT "Bacteriophage T4 genome analysis."  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=91203817; PubMed=2017138;  
 RA Ganz A., Kruse U., Rueger W.;  
 RT "Gene product dsbA of bacteriophage T4 binds to late promoters and  
 enhances late transcription."  
 RL Mol. Gen. Genet. 225:427-433 (1991).  
 CC -!- FUNCTION: BINDS DOUBLE-STRANDED DNA; INTERACTS PREFERENTIALLY WITH  
 T4 LATE PROMOTER REGIONS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X15818; CAA33813.1; -.  
 DR EMBL; AF158101; AAD42567.1; -.  
 DR PIR; S05557; S05557.  
 KW DNA replication; DNA-binding.  
 SQ SEQUENCE 89 AA; 10379 MW; 0152649899E09319 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 20; DB 1; Length 89;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
 ||||  
 Db 47 KDEL 50

Search completed: August 26, 2003, 16:06:21  
 Job time : 2.99115 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: August 26, 2003, 16:04:38 ; Search time 4.31858 Seconds  
(without alignments)  
239.016 Million cell updates/sec

Title: US-09-696-872-37

Perfect score: 20

Sequence: 1 KDEL 4

Scoring table: BLOSUM62  
Gapox 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_todent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	21	11 Q9ESX0	Q9ESX0 mus musculu
2	20	100.0	35	10 Q9S8D1	Q9S8D1 gymnaema syl
3	20	100.0	40	4 Q13833	Q13833 homo sapien
4	20	100.0	40	4 Q13832	Q13832 homo sapien
5	20	100.0	41	2 O06527	O06527 lactobacilli
6	20	100.0	41	11 Q9ECR6	Q9ECR6 mus musculu
7	20	100.0	41	16 Q8DRV6	Q8DRV6 streptococc
8	20	100.0	42	5 Q9NG47	Q9NG47 aedes albp
9	20	100.0	46	2 Q57372	Q57372 vibrio chol
10	20	100.0	48	4 Q9UDJ7	Q9UDJ7 homo sapien
11	20	100.0	49	16 Q8EB07	Q8EB07 shewanella
12	20	100.0	50	2 Q50051	Q50051 mycobacteri
13	20	100.0	50	5 Q8STJ8	Q8STJ8 anartia fat
14	20	100.0	50	5 Q9SVF9	Q9SVF9 dryas iulia
15	20	100.0	50	16 Q8YKY6	Q8YKY6 anabaena sp
16	20	100.0	51	5 Q8T5E2	Q8T5E2 anartia lyc

17	20	100.0	51	5 Q95VG2	Q95VG2 eueides ali
18	20	100.0	51	5 Q95VG0	Q95VG0 eueides vib
19	20	100.0	51	5 Q8T5D0	Q8T5D0 anartia fat
20	20	100.0	52	5 Q8T5D9	Q8T5D9 anartia ama
21	20	100.0	52	5 Q8T5D1	Q8T5D1 anartia fat
22	20	100.0	52	10 Q9SFX4	Q9SFX4 picea adies
23	20	100.0	52	10 Q65057	Q65057 picea maria
24	20	100.0	53	5 Q8T5E1	Q8T5E1 anartia chr
25	20	100.0	53	7 Q30456	Q30456 equus cabal
26	20	100.0	53	7 Q30476	Q30476 equus cabal
27	20	100.0	53	7 Q30471	Q30471 equus cabal
28	20	100.0	53	7 Q30470	Q30470 equus cabal
29	20	100.0	53	7 Q9BCX3	Q9BCX3 equus cabal
30	20	100.0	53	10 Q9SPX3	Q9SPX3 picea abies
31	20	100.0	53	16 Q8KET6	Q8KET6 chlorobium
32	20	100.0	54	10 Q9XIA7	Q9XIA7 arabidopsis
33	20	100.0	57	11 Q9ES41	Q9ES41 rattus norv
34	20	100.0	58	16 Q8U9H6	Q8U9H6 agrobacteri
35	20	100.0	59	5 Q8T5E0	Q8T5E0 anartia chr
36	20	100.0	60	16 Q8YNB4	Q8YNB4 anabaena sp
37	20	100.0	60	17 Q97C12	Q97C12 thermoplasm
38	20	100.0	61	11 Q88367	Q88367 mus musculu
39	20	100.0	61	16 Q9ADC0	Q9ADC0 streptomyce
40	20	100.0	62	13 Q91801	Q91801 xenopus lae
41	20	100.0	62	13 Q8JH23	Q8JH23 ostracion s
42	20	100.0	63	10 Q9AX34	Q9AX34 oryza sativ
43	20	100.0	64	4 Q96GN8	Q96GN8 homo sapien
44	20	100.0	65	5 Q8T5D2	Q8T5D2 anartia fat
45	20	100.0	65	5 Q8T5D8	Q8T5D8 anartia ama

## ALIGNMENTS

### RESULT 1

ID Q9ESX0 PRELIMINARY; PRT; 21 AA.  
AC Q9ESX0;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Gephyrin (Fragment).  
GN GPHN OR GEPHYRIN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OC NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20420367; PubMed=10963686;  
RA Ramming M., Kins S., Werner N., Hermann A., Betz H., Kirsch J.;  
RT "Diversity and phylogeny of gephyrin: Tissue-specific splice variants,  
RT gene structure, and sequence similarities to molybdenum cofactor-  
RT synthesizing and cytoskeleton-associated proteins.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:10266-10271(2000).  
DR EMBL: AJ278768; CAC06105.1; -  
DR MGD; MGI:109602; Gphn.  
FT NON\_TER  
FT NON\_TER  
SQ SEQUENCE 21 AA; 2627 MW; 0820F760BC776F9A CRC64;

Query Match 100.0%; Score 20; DB 11; Length 21;  
Best local similarity 100.0%; Pred. NO. 3; e+02;  
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDEL 4  
|||  
Db 9 KDEL 12

RESULT 2  
Q9S8D1 PRELIMINARY; PRT; 35 AA.

```
AC 09SBD1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE GURMARIN=SWETNESS-suppressing polypeptide.
OS Gymnema sylvestre.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Gentianales; Apocynaceae; Periplocoidae; Gymnema.
OC NCBI_TaxId=4068;
RN [1]
RN SEQUENCE.
RX MEDLINE=96068945; PubMed=8534991;
RA Oca M., Ariyoshi Y.;
RT "Location of the disulfide bonds of the sweetness-suppressing
RT polypeptide gurmarin.";
RL Bioact. Biotechnol. Biochem. 59:1956-1957(1995).
DR HSP; P25810; IGUR.
SQ SEQUENCE 35 AA; 4233 MW; 661BCAE66E904CB6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 10; Length 35;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 6 KDEL 9

RESULT 3
Q13833 PRELIMINARY; PRT; 40 AA.
AC 013833;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE B2 bradykinin receptor basal promoter; allele Bp-58-T (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=96209920; PubMed=8655154;
RA Braun A., Maier E., Kammerer S., Mueller B., Roscher A.A.;
RT "A novel sequence polymorphism in the promoter region of the human
RT bradykinin B2-receptor gene.";
RL Hum. Genet. 97:688-689(1996).
DR EMBL; X91664; CAA62852.1; -.
KW Receptor.
FT NON_TER 1 1
FT VARIANT 18 19 IT -> XS.
FT NON_TER 40 40
SQ SEQUENCE 40 AA; 4152 MW; 1408B9AD371EB17F CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 4; Length 40;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 23 KDEL 26

RESULT 4
Q13832 PRELIMINARY; PRT; 40 AA.
AC 013832;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE B2 bradykinin receptor basal promoter; allele Bp-58-C (Fragment).
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=96209920; PubMed=8655154;
RA Braun A., Maier E., Kammerer S., Mueller B., Roscher A.A.;
RT "A novel sequence polymorphism in the promoter region of the human
RT bradykinin B2-receptor gene.";
RL Hum. Genet. 97:688-689(1996).
DR EMBL; X91663; CAA62851.1; -.
KW Receptor.
FT NON_TER 1 1
FT VARIANT 18 19 TT -> XS.
FT NON_TER 40 40
SQ SEQUENCE 40 AA; 4140 MW; 3908B9AD371EF4A5 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 4; Length 40;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 23 KDEL 26

RESULT 5
Q06527 PRELIMINARY; PRT; 41 AA.
ID 006527;
AC 006527;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Cystathionine gamma-lyase homolog (Fragment).
OS Lactobacillus fermentum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OC NCBI_TaxId=1613;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=BR11;
RX MEDLINE=97294473; PubMed=9150229;
RA Turner M.S., Timms P., Hafner L.M., Giffard P.M.;
RT "Identification and characterization of a basic cell surface-located
RT protein from Lactobacillus fermentum BR11.";
RL J. Bacteriol. 179:3310-3316(1997).
RN [2]
RN SEQUENCE FROM N.A.
RX STRAIN=BR11;
RX MEDLINE=99194729;
RA Turner M.S., Woodberry T., Hafner L.M., Giffard P.M.;
RT "The bcpA locus of Lactobacillus fermentum BR11 encodes an L-cysteine
RT uptake system.";
RL J. Bacteriol. 181:2192-2198(1999).
DR EMBL; U97348; AAC45329.1; -.
DR InterPro; IPR000277; Cys_Met_Meta_PP.
DR Pfam; PF01053; Cys_Met_Meta_PP; 1.
KW Lyase.
FT NON_TER 1 1
FT SEQUENCE 41 AA; 4647 MW; 7875F22817910FB7 CRC64;

Query Match
Best Local Similarity 100.0%; Score 20; DB 2; Length 41;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
DB 13 KDEL 16

RESULT 6
Q9EQE6 PRELIMINARY; PRT; 41 AA.
ID Q9EQE6
```



```

AC Q9E0E6;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE General transcription factor II-1 repeat domain protein 1
   (Fragment).
GN GTF2IRD1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Sv;
MEDLINE=21250999; PubMed=11352562;
RA Duxkin M.E., Keck-Waggoner C.L., Popescu N.C., Thorgeirsson S.S.;
RT "Integration of a c-myc Transgene Results in Disruption of the Mouse
Gtf2ird1 Gene, the Homologue of the Human GTF2IRD1 Gene Hemizyously
Deleted in Williams-Beuren Syndrome.";
RL Genomics 73:20-27(2001).
DR EMBL; AF257477; AAG44656.1; -.
DR MGI; MGI:1861942; Gtf2ird1.
FT NON_TER
SQ SEQUENCE 41 AA; 4482 MW; D7E29ECA570059DC CRC64;

Query Match 100.0%; Score 20; DB 11; Length 41;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
   |||
   |||
Db 27 KDEL 30

RESULT 7
Q8DRV6 PRELIMINARY; PRT; 41 AA.
AC Q8DRV6;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN SMW.2105.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AE015032; AAN59699.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 41 AA; 4849 MW; CD8F7DC062E6D60B CRC64;

Query Match 100.0%; Score 20; DB 16; Length 41;
Best Local Similarity 100.0%; Pred. No. 7.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
   |||
   |||
Db 5 KDEL 8

RESULT 8
Q9NG47 PRELIMINARY; PRT; 42 AA.

```

```

AC Q9NG47;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Ribosomal S5 protein (Fragment).
OS Aedes albopictus (Forest day mosquito).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes.
OX NCBI_TaxID=7160;
RN [1]
RP SEQUENCE FROM N.A.
RA Yang F.F., Zhao T.T., Li D.D.;
RT "Differentially expressed genes of Aedes albopictus orally infected
RT with dengue-2 virus.";
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF263471; AAF73440.1; -.
DR InterPro; IPR000235; Ribosomal_S7; 1.
DR Pfam; PF00177; Ribosomal_S7; 1.
DR ProDom; PD000817; Ribosomal_S7; 1.
FT NON_TER
SQ SEQUENCE 42 AA; 4711 MW; CE1FB5123BE053A2 CRC64;

Query Match 100.0%; Score 20; DB 5; Length 42;
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
   |||
   |||
Db 31 KDEL 34

RESULT 9
O57372 PRELIMINARY; PRT; 46 AA.
AC O57372;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Cholera toxin A2.
GN CTX42.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrrio cholerae.
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4260B;
RX MEDLINE=94237453; PubMed=8181723;
RA LeDens M., Holmgren J.;
RT "Structure and arrangement of the cholera toxin genes in Vibrio
RT cholerae O139.";
RL FEMS Microbiol. Lett. 117:197-202(1994).
DR EMBL; X76391; CAAS3975.1; -.
DR EMBL; X76390; CAAS3974.1; -.
DR HSSP; P01555; IXTC.
DR InterPro; IPR001144; Enterotoxin_A.
DR InterPro; IPR000886; ER_target.
DR Pfam; PF01375; Enterotoxin_A; 1.
DR ProSite; PS00014; ER_TARGET; 1.
SQ SEQUENCE 46 AA; 5447 MW; 1B6085A02E8889D6 CRC64;

Query Match 100.0%; Score 20; DB 2; Length 46;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDEL 4
   |||
   |||
Db 43 KDEL 46

RESULT 10
Q9UDJ7 PRELIMINARY; PRT; 48 AA.

```

AC Q9UDJ7; 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE Nucleolar shuttle protein B-23-38 kDa major NOS-binding protein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=93300776; PubMed=8314759;  
 RA Adachi Y., Copeland T.D., Hatanaka M., Oroszlan S.;  
 RT "Nucleolar targeting signal of Rex protein of human T-cell leukemia virus type I specifically binds to nucleolar shuttle protein B-23.";  
 RL J. Biol. Chem. 268:13930-13934(1993).  
 FT NON\_TER 1  
 FT NON\_CONS 18 19  
 FT NON\_CONS 33 34  
 FT NON\_TER 48 48  
 SQ SEQUENCE 48 AA; 4783 MW; 5D7D533BD386A9B CRC64;

Query Match 100.0%; Score 20; DB 4; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 8.5e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
 DB 5 KDEL 8

## RESULT 11

O8EB07 PRELIMINARY; PRT; 49 AA.  
 AC Q8EB07;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 CN S03729.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadaceae; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RX MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C., Reed T.D., Eisen J.A., Seshadri R., Ward N., Mehta B., Clayton R.A., Meyer R.T., Tsai A., Scott J., Beanan M., Binkac L., Daugherty S., DeWay R.T., Dodson R.J., Durkin A.S., Haft D.H., Klonay J.F., Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M., Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C., Mueller J., Kouri H., Gill J., Uterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;  
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium Shewanella oneidensis.";  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL; AE015806; AAN56713.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 49 AA; 5340 MW; 7ABEB48EAD0FFA80 CRC64;

Query Match 100.0%; Score 20; DB 16; Length 49;  
 Best Local Similarity 100.0%; Pred. No. 8.7e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
 DB 31 KDEL 34

RESULT 12  
 O50051 PRELIMINARY; PRT; 50 AA.  
 AC Q50051;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last annotation update)  
 DE U2266p.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Smith D.R.;  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U15182; AAA62966.1; -.  
 SQ SEQUENCE 50 AA; 5563 MW; 5B5A4419AD5DD6C2 CRC64;

Query Match 100.0%; Score 20; DB 2; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
 DB 41 KDEL 44

## RESULT 13

O8ST38 PRELIMINARY; PRT; 50 AA.  
 AC Q8ST38;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Triosephosphate isomerase (Fragment).  
 OS Anartia fatima.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Papilionoidea; Nymphalidae; Nymphalinae; Anartia.  
 OX NCBI\_TaxID=145889;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Blum M.J., Bermingham E., Dasmanapetra K.;  
 RT "A Molecular Phylogeny of the Neotropical Butterfly Genus Anartia (Lepidoptera: Nymphalidae).";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY038684; AAM03332.1; -.  
 DR EMBL; AY038685; AAM03333.1; -.  
 DR InterPro; IPR000652; Triophos\_ismrse.  
 DR Pfam; PF00121; TIM; 1.  
 DR ProDom; PD001005; Triophos\_ismrse; 1.  
 KW Isomerase.  
 FT NON\_TER 1  
 FT NON\_TER 50  
 FT NON\_TER 50  
 SQ SEQUENCE 50 AA; 5450 MW; F985EA898F5A3BAF CRC64;

Query Match 100.0%; Score 20; DB 5; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KDEL 4  
 DB 2 KDEL 5

## RESULT 14

O95VF9

Search completed: August 26, 2003, 16:09:20  
 Job time : 7.31858 secs

```

ID Q95VF9 PRELIMINARY; PRT; 50 AA.
AC Q95VF9;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Triosephosphate isomerase (Fragment).
GN TPI.
OS Dryas julia (julia butterfly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Papilionoidea; Nymphalidae; Heliconiinae; Dryas.
OX NCBI_TaxID=33453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STRI-B-293-Tpi-1;
RA Beltran M.S., Jiggins C.D., Bull V., McMillan O., Bermingham E.,
RA Mallet J., Linares M.;
RT "Phylogenetic discordance at the species boundary: comparative gene
RL genealogies between Heliconius butterflies.";
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF413797; AAL08558.1; -.
DR InterPro; IPR00652; Triophos_ismrse.
DR Pfam; PF00121; TIM; 1.
DR ProDom; PD001005; Triophos_ismrse; 1.
KW isomerase.
FT NON_TER 1
FT NON_TER 50
SQ SEQUENCE 50 AA; 5394 MW; AF3E2F54436BAFF8 CRC64;

```

Query Match 100.0%; Score 20; DB 5; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 KDEL 4
    |||
    |||
Db 3 KDEL 6

```

## RESULT 15

```

ID Q8YKY6 PRELIMINARY; PRT; 50 AA.
AC Q8YKY6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Transposase.
GN ASR7152.
OS Anabaena sp. (strain PCC 7120).
OG Plasmid pCC7120alpha.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21595285; PubMed=11759840;
RA Kaneo T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003600; BAB78236.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 50 AA; 5760 MW; BE163ACA7B5C489B CRC64;

```

Query Match 100.0%; Score 20; DB 16; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;  
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 KDEL 4
    |||
    |||
Db 8 KDEL 11

```

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_p2n model

Run on: August 26, 2003, 16:08:15 ; Search time 177.416 Seconds

(without alignments)  
922.344 Million cell updates/sec

Title: US-09-696-872-37

Perfect score: 20

Sequence: 1 XDEL 4

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09696872/runat\_26082003\_151137\_3223/app\_query.fasta\_1.462  
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09696872.OCGN\_1.1.3608.0@runat\_26082003\_151137\_3223 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_man:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	12	6	BD174826
2	20	100.0	16	6	AX710937
3	20	100.0	16	6	BD001078
4	20	100.0	16	6	BD001507
5	20	100.0	17	6	AX215210
6	20	100.0	17	6	AX215211
7	20	100.0	17	6	AX215212
8	20	100.0	17	6	AX216059
9	20	100.0	19	6	A58382
10	20	100.0	20	6	AR160690
11	20	100.0	20	6	AR183981
12	20	100.0	20	6	AR221427
13	20	100.0	20	6	AR221428
14	20	100.0	20	6	AR221429
15	20	100.0	20	6	BD138303
16	20	100.0	20	6	BD138304
17	20	100.0	21	6	BD143817
18	20	100.0	21	6	I26155
19	20	100.0	21	6	I86398
20	20	100.0	22	6	AR135254
21	20	100.0	23	6	AX343349
22	20	100.0	23	6	AX343350
23	20	100.0	23	6	BD133175
24	20	100.0	24	6	BD133176
25	20	100.0	24	6	A10287
26	20	100.0	24	6	A10288
27	20	100.0	24	6	A10288
28	20	100.0	24	6	A10288
29	20	100.0	24	6	A16561
30	20	100.0	24	6	A16562
31	20	100.0	24	6	A16562
32	20	100.0	24	6	A16562
33	20	100.0	24	6	A33707
34	20	100.0	24	6	A33707
35	20	100.0	24	6	A33708
36	20	100.0	24	6	A33708
37	20	100.0	24	6	A82182
38	20	100.0	24	6	AR307631
39	20	100.0	24	6	AR307641
40	20	100.0	24	6	BD132633
41	20	100.0	24	6	BD174866
42	20	100.0	25	6	AR137362
43	20	100.0	25	6	AX037484
44	20	100.0	25	6	AX370858
45	20	100.0	25	6	BD075236

RESULT 1

#### ALIGNMENTS



Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 6  
Gaps: 0

US-09-696-872-37 (1-4) x BD001078 (1-16)

QY 1 LysApGluLeu 4  
2 AAGATGAACCTT 13

RESULT 4  
LOCUS BD001507  
DEFINITION Method and reagent for inhibiting viral replication.  
ACCESSION BD001507  
VERSION BD001507.1 GI:18626066  
KEYWORDS JP 2000342286-A/238.  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1 (bases 1 to 16)  
AUTHORS Draper,K.G., Dadykiz,L.W., Macswigen,J.A., Maysejak,D.G.,  
TITLE Holesch,J.J. and Mamone,A.J.  
METHOD Method and reagent for inhibiting viral replication  
JOURNAL Patent: JP 2000342286-A 238 12-DEC-2000;  
RIBOZYME PHARMACEUTICALS INC

COMMENT OS Artificial Sequence  
PN JP 2000342286-A/238  
PD 12-DEC-2000  
PR 11-MAY-1992 US 07/882689, 14-MAY-1992 US 07/882712 PR  
14-MAY-1992 US 07/882713, 14-MAY-1992 US 07/882714 PR  
14-MAY-1992 US 07/882822, 14-MAY-1992 US 07/882824 PR  
14-MAY-1992 US 07/882886, 14-MAY-1992 US 07/882888 PR  
14-MAY-1992 US 07/882889, 14-MAY-1992 US 07/882921 PR  
14-MAY-1992 US 07/882922, 14-MAY-1992 US 07/883823 PR  
14-MAY-1992 US 07/883849, 14-MAY-1992 US 07/884073 PR  
14-MAY-1992 US 07/884074, 14-MAY-1992 US 07/884333 PR  
14-MAY-1992 US 07/884422, 14-MAY-1992 US 07/884431 PR  
14-MAY-1992 US 07/884436, 14-MAY-1992 US 07/884521 PR  
31-JUL-1992 US 07/923738, 26-AUG-1992 US 07/935854 PR  
26-AUG-1992 US 07/936086, 18-SEP-1992 US 07/948359 PR  
15-OCT-1992 US 07/963322, 07-DEC-1992 US 07/987129 PR  
07-DEC-1992 US 07/987130, 07-DEC-1992 US 07/987133 PR  
KENNETH G DRAPER, LEC W DADYKIZ, JAMES A MACSWIGEN, PI DENNIS G  
WAYSEJAK,

PI JAMES J HOLESEK, ANTHONY J MAMONE  
PC C12N15/09, C12N5/10, C12N7/00//A61K38/43, A61K39/125, A61K39/13,  
PC A61K39/135,  
PC A61K39/145, A61K39/21, A61K39/23, A61K39/245, A61K39/29, A61K46/00,  
PC A61P1/16,  
PC A61P31/14, A61P31/16, A61P31/18, A61P31/22, A61P35/02, C12Q1/68, PC  
(C12N15/09, C12R1:93), C12N15/00, C12N5/00, A61K37/48, (C12N15/00, PC  
C12R1:93)

CC Key location/Qualifiers  
FH source 1.16 location/Qualifiers  
FT source 1.16 location/Qualifiers

FEATURES  
source 1.16 location/Qualifiers  
1.16 /organism="synthetic construct"  
/mol\_type="genomic RNA"  
/db\_xref="taxon:32630"

BASE COUNT 7 a 1 c 5 g 3 t

ORIGIN

Alignment Scores:  
Pred. No.: 158 Length: 16  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservatve: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x BD001507 (1-16)

QY 1 LysApGluLeu 4  
2 AAGATGAACCTT 13

RESULT 5  
LOCUS AX215210/c  
DEFINITION Sequence 652 from Patent WO0159103.  
ACCESSION AX215210  
VERSION AX215210.1 GI:15525253  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Blatt,L., Mcswigen,J. and Chowitra,B.M.  
TITLE Method and reagent for the modulation and diagnosis of cd20 and  
JOURNAL nogo gene expression  
PATENT: WO 0159103-A 652 16-AUG-2001;  
RIBOZYME PHARMACEUTICALS, INC. (US) ; Blatt, Lawrence (US) ;  
Mcswigen, James (US) ; Chowitra, Bharat M. (US)

FEATURES  
source 1.17 location/Qualifiers  
1.17 /organism="synthetic construct"  
/mol\_type="mRNA"  
/db\_xref="taxon:32630"  
/note="Nucleic Acid"

BASE COUNT 4 a 2 c 4 g 7 t

ORIGIN

Alignment Scores:

Pred. No.: 168 Length: 17  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservatve: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0

US-09-696-872-37 (1-4) x AX215210 (1-17)

QY 1 LysApGluLeu 4  
16 AAGATGAACCTC 5

RESULT 6  
LOCUS AX215211/c  
DEFINITION Sequence 653 from Patent WO0159103.  
ACCESSION AX215211  
VERSION AX215211.1 GI:15525254  
KEYWORDS  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Blatt,L., Mcswigen,J. and Chowitra,B.M.  
TITLE Method and reagent for the modulation and diagnosis of cd20 and  
JOURNAL nogo gene expression  
PATENT: WO 0159103-A 653 16-AUG-2001;  
RIBOZYME PHARMACEUTICALS, INC. (US) ; Blatt, Lawrence (US) ;  
Mcswigen, James (US) ; Chowitra, Bharat M. (US)

FEATURES  
source 1.17 location/Qualifiers  
1.17 /organism="synthetic construct"  
/mol\_type="mRNA"  
/db\_xref="taxon:32630"  
/note="Nucleic Acid"

BASE COUNT 5 a 2 c 3 g 7 t

ORIGIN

Alignment Scores:

Pred. No.:	168	Length:	17
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-696-872-37 (1-4) x AX215211 (1-17)

QY 1 LysaepgJlu4

DB 15 AAAGATGAATC 4

RESULT 7

LOCUS AX215212/c

DEFINITION Sequence 654 from Patent WO0159103.

ACCESSION AX215212

VERSION AX215212.1 GI:15525255

KEYWORDS

SOURCE synthetic construct

ORGANISM artificial sequences.

REFERENCE 1 Blatt, L., McSwigen, J. and Chowrita, B.M. Method and reagent for the modulation and diagnosis of cd20 and nogo gene expression

AUTHORS Patent: WO 0159103-A 654 16-AUG-2001;

JOURNAL RIBOZYME PHARMACEUTICALS, INC. (US); Blatt, Lawrence (US); McSwigen, James (US); Chowrita, Bharat M. (US)

FEATURES

source 1..17

/organism="synthetic construct"

/mol\_type="mRNA"

/db\_xref="taxon:32630"

/note="Nucleic Acid"

BASE COUNT 5 a 2 c 4 g 6 t

ORIGIN

Alignment Scores:

Pred. No.:	168	Length:	17
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-696-872-37 (1-4) x AX215212 (1-17)

QY 1 LysaepgJlu4

DB 12 AAAGATGAATC 1

RESULT 8

LOCUS AX216059/c

DEFINITION Sequence 1501 from Patent WO0159103.

ACCESSION AX216059

VERSION AX216059.1 GI:15526102

KEYWORDS

SOURCE synthetic construct

ORGANISM artificial sequences.

REFERENCE 1 Blatt, L., McSwigen, J. and Chowrita, B.M. Method and reagent for the modulation and diagnosis of cd20 and nogo gene expression

AUTHORS Patent: WO 0159103-A 1501 16-AUG-2001;

JOURNAL RIBOZYME PHARMACEUTICALS, INC. (US); Blatt, Lawrence (US); McSwigen, James (US); Chowrita, Bharat M. (US)

FEATURES

Location/Qualifiers

source 1..17

/organism="synthetic construct"

/mol\_type="mRNA"

/db\_xref="taxon:32630"

/note="Nucleic Acid"

BASE COUNT 6 a 2 c 3 g 6 t

ORIGIN

Alignment Scores:

Pred. No.:	168	Length:	17
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-696-872-37 (1-4) x AX216059 (1-17)

QY 1 LysaepgJlu4

DB 14 AAAGATGAATC 3

RESULT 9

LOCUS A58382

DEFINITION Sequence 11 from Patent WO9637607.

ACCESSION A58382

VERSION A58382.1 GI:3714038

KEYWORDS

SOURCE unidentified

ORGANISM unidentified

REFERENCE 1 Braun, A. and Kammerer, S. NUCLEIC ACID WITH A POLYMORPHIC DOMAIN OF A B2 BRADYKININ RECEPTOR

AUTHORS GEN Patent: WO 9637607-A 11 28-NOV-1996;

JOURNAL ROSCHER ADELBERT (DE)

COMMENT Other publication DE 19518931 960919.

FEATURES

source 1..19

/organism="unidentified"

/mol\_type="genomic DNA"

/db\_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 4 t

ORIGIN

Alignment Scores:

Pred. No.:	189	Length:	19
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-696-872-37 (1-4) x A58382 (1-19)

QY 1 LysaepgJlu4

DB 8 AAAGATGAATC 19

RESULT 10

LOCUS ARI60690

DEFINITION Sequence 16 from patent US 6255105.

ACCESSION ARI60690

VERSION ARI60690.1 GI:16225163

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 20) Marchetti, A., Buttitta, F., Smith, G.H. and Callahan, R.





Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x AR221429 (1-20)

Qy 1 LysaSPGjLeu 4  
 |||||  
 Db 19 AAGGACGAGCTG 8

## RESULT 15

BD138303 20 bp DNA linear PAT 18-SEP-2002  
 LOCUS Antisense modulation of human MDM2 expression.

DEFINITION BD138303

ACCESSION BD138303.1 GI:23233248

VERSION JP 2002508944-A/229.

KEYWORDS unclassified

SOURCE unclassified

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 20)

AUTHORS Miraglia,L.J., Nero,P., Graham,M.J., Monia,B.P. and Cowseert,L.M.

TITLE Antisense modulation of human MDM2 expression

JOURNAL Patent: JP 2002508944-A 229 26-MAR-2002;

ISIS PHARMACEUTICALS INC

OS Unidentified

PN JP 2002508944-A/229

PD 26-MAR-2002 JP 2000538025

PF 26-MAR-1999 JP 2000538025

PR 26-MAR-1998 US 09/048810

PI LOREN J MIRAGLIA, PAMELA NERO, MARK J GRAHAM, BRETT P MONIA, LEX M

## COMMENT

PI COMSERT  
 PC C12N15/09,A61K48/00,A61P9/10,A61P17/06,A61P35/00,C07H21/04//  
 PC C12Q1/68,  
 PC C12N15/00  
 CC Strandedness: Single;  
 CC Topology: Linear;  
 CC Antisense modulation of human MDM2 expression FH Key  
 CC Location/Qualifiers  
 FT source 1..20  
 Location/Qualifiers

FEATURES  
 source 1..20  
 /organism="Unidentified".  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:32644"

BASE COUNT 8 a 3 c 4 g 5 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 199 Length: 20  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-696-872-37 (1-4) x BD138303 (1-20)

Qy 1 LysaSPGjLeu 4  
 |||||  
 Db 2 AAGGATGAGCTA 13

Search completed: August 26, 2003, 18:30:19  
 Job time : 179.416 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_p1ue\_p2n model

Run on: August 26, 2003, 16:04:40 ; Search time 13.6991 Seconds  
(without alignments)  
788.208 Million cell updates/sec

Title: US-09-696-872-37  
Perfect score: 20  
Sequence: 1 KDEL 4

Scoring table:  
BIOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 segs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

MODEL=frame\_p2n.model -DEV=xlh  
-O=/cgnr\_1/USPTO.spool/US09696872/runat\_26082003\_151137\_3213/app\_query.fasta.1.462  
-DB=N-Geneseq\_19Jun03 -QPMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=PCCT -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=PCO -NCM=score -HEAPSIZE=500 -MITLEN=0 -MAXLEN=2000000000  
-USER=US09696872.@CGN\_1.1.0.@runat\_26082003\_151137\_3213 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGESQDENTRY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N-Geneseq\_19Jun03:\*  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	12	AA235104	DNA encoding endop
2	20	100.0	12	AA235104	Endoplasmic reticu
3	20	100.0	12	AA235104	Anti-human calisiv
4	20	100.0	12	AA235104	Oligonucleotide SE
5	20	100.0	12	AA235104	Oligonucleotide SE
6	20	100.0	12	AA235104	Oligonucleotide SE
7	20	100.0	12	AA235104	Oligonucleotide SE
8	20	100.0	12	AA235104	Oligonucleotide SE
9	20	100.0	12	AA235104	Oligonucleotide SE
10	20	100.0	12	AA235104	Oligonucleotide SE
11	20	100.0	12	AA235104	Oligonucleotide SE
12	20	100.0	12	AA235104	Oligonucleotide SE
13	20	100.0	12	AA235104	Oligonucleotide SE
14	20	100.0	12	AA235104	Oligonucleotide SE
15	20	100.0	12	AA235104	Oligonucleotide SE
16	20	100.0	12	AA235104	Oligonucleotide SE
17	20	100.0	12	AA235104	Oligonucleotide SE
18	20	100.0	12	AA235104	Oligonucleotide SE
19	20	100.0	12	AA235104	Oligonucleotide SE
20	20	100.0	12	AA235104	Oligonucleotide SE
21	20	100.0	12	AA235104	Oligonucleotide SE
22	20	100.0	12	AA235104	Oligonucleotide SE
23	20	100.0	12	AA235104	Oligonucleotide SE
24	20	100.0	12	AA235104	Oligonucleotide SE
25	20	100.0	12	AA235104	Oligonucleotide SE
26	20	100.0	12	AA235104	Oligonucleotide SE
27	20	100.0	12	AA235104	Oligonucleotide SE
28	20	100.0	12	AA235104	Oligonucleotide SE
29	20	100.0	12	AA235104	Oligonucleotide SE
30	20	100.0	12	AA235104	Oligonucleotide SE
31	20	100.0	12	AA235104	Oligonucleotide SE
32	20	100.0	12	AA235104	Oligonucleotide SE
33	20	100.0	12	AA235104	Oligonucleotide SE
34	20	100.0	12	AA235104	Oligonucleotide SE
35	20	100.0	12	AA235104	Oligonucleotide SE
36	20	100.0	12	AA235104	Oligonucleotide SE
37	20	100.0	12	AA235104	Oligonucleotide SE
38	20	100.0	12	AA235104	Oligonucleotide SE
39	20	100.0	12	AA235104	Oligonucleotide SE
40	20	100.0	12	AA235104	Oligonucleotide SE
41	20	100.0	12	AA235104	Oligonucleotide SE
42	20	100.0	12	AA235104	Oligonucleotide SE
43	20	100.0	12	AA235104	Oligonucleotide SE
44	20	100.0	12	AA235104	Oligonucleotide SE
45	20	100.0	12	AA235104	Oligonucleotide SE

## ALIGNMENTS

RESULT 1  
AA235104  
ID AA235104 standard; DNA, 12 BP.  
AA235104;  
13-MAR-2000 (first entry)  
DNA encoding endoplasmic reticulum retention signal.  
Endoplasmic reticulum; monocotyledonous plant;  
Monocot; rice; wheat; transgenic plant; mammalian polypeptide;  
antibody; expression cassette; ss.  
Synthetic.  
MO9966026-A2.

XX 23-DEC-1999.  
 PD 15-JUN-1999; 99WO-US13584.  
 XX PF 15-JUN-1998; 98US-0089322.  
 XX PR 15-JUN-1998; 98US-0089322.  
 XX PA (INNE-) INNES CENT JOHN.  
 XX PA (CHRI/) CHRISTOU P.  
 PI Christou P, Stroger E, Fischer R, Martin-Vaquero C, Schillberg S;  
 PI Ma JK;  
 XX WPI; 2000-097739/08.  
 DR P-PSDB; AAY32395.  
 XX  
 PT Production of mammalian polypeptides, especially antibodies, by  
 PT monocotyledonous plants -  
 XX  
 XS Disclosure; Page 5; 76pp; English.

This DNA sequence codes for an endoplasmic reticulum (ER) retention  
 CC signal (see AAY32395). Levels of mammalian polypeptide expression in  
 CC transgenic monocotyledonous plants can be enhanced by use of this  
 CC signal. The invention provides rice, wheat and other monocot plants  
 CC that have been transformed with expression cassettes for production  
 CC of mammalian polypeptides, such as antibodies. ER retention signals  
 CC (see AAY32395-96), 5' untranslated regions and leader peptides are  
 CC employed in various combinations to provide high expression yield.  
 CC Plant cell or seed containing polypeptides selected from Fv, Fab,  
 CC F(ab)2, diabody, dimeric scfv, whole antibody and four-chain  
 CC secretory antibody are claimed. Multi-chain complexes such as  
 CC four-chain secretory antibodies are produced by expression of  
 CC component polypeptides from separate vectors, all introduced into  
 CC the same cell by transformation.

SEQ Sequence 12 BP; 5 A; 2 C; 3 G; 2 T; 0 other;

#### Alignment Scores:

Pred. No.:	138	Length:	12
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-696-872-37 (1-4) x AAZ35104 (1-12)

QY 1 LysApGjuleu 4

DB 1 AAAGATGAGCTC 12

#### RESULT 2

ID AAF57883 standard; DNA; 12 BP.

AC AAF57883;

DT 20-APR-2001 (first entry)

DE Endoplasmic reticulum retention signal peptide #1 coding sequence.

KW Endoplasmic reticulum; ER; retention signal; transgenic plant;

KW expression cassette; ds.

OS Unidentified.

PN WO200105936-A2.

PD 25-JAN-2001.

PF 18-JUL-2000; 2000WO-US19721.

PR 19-JUL-1999; 99US-0144513.  
 PR 07-JUL-2000; 2000US-0611736.  
 XX  
 XX (INNE-) INNES CENT JOHN.  
 XX PA  
 XX PI Christou P, Kohli A;  
 XX PI  
 XX DR WPI; 2001-147332/15.  
 DR P-PSDB; AAB66569.  
 XX

PT Producing population of transgenic plants involves transforming plant  
 PT cell with minimal transgene expression cassette by direct DNA transfer  
 PT techniques and regenerating transgenic plants from transformed cells -  
 XX  
 PS Disclosure; Page 10; 42pp; English.

CC The present invention relates to a method for producing a population of  
 CC transgenic plants, comprising transforming intact plant cells with  
 CC minimal transgene expression cassettes by direct DNA transfer techniques,  
 CC and regenerating a population of transgenic plants from transformed plant  
 CC cells. The present sequence is the coding sequence for an endoplasmic  
 CC reticulum (ER) retention signal peptide. This sequence can be used in the  
 CC expression cassette of the present invention.

SEQ Sequence 12 BP; 5 A; 2 C; 3 G; 2 T; 0 other;

#### Alignment Scores:

Pred. No.:	138	Length:	12
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-09-696-872-37 (1-4) x AAF57883 (1-12)

QY 1 LysApGjuleu 4

DB 1 AAAGATGAGCTC 12

#### RESULT 3

ID ABT15850 standard; DNA; 12 BP.

AC ABT15850;

DT 28-MAR-2003 (first entry)

DE Anti-human calicivirus monoclonal antibody related DNA SEQ ID No 24.

KW Transgenic plant; anti-human calicivirus monoclonal antibody; gene; ds.

OS Unidentified.

PN JP2002253262-A.

PD 10-SEP-2002.

PE 05-MAR-2001; 2001JP-0060462.

PR 05-MAR-2001; 2001JP-0060462.

PA (FRON-) FRONTIER SCI KK.

PA (HOKK-) HOKKAIDO GREEN BIO KENKYUSHO KK.

PA (SUGI/) SUGIMOTO C.

PA (UEDA/) UEDA I.

DR WPI; 2003-150966/15.

DR P-PSDB; ABJ19274.

PT A new transgenic plant cell useful for producing anti-human calicivirus  
 PT monoclonal antibody -

XX

Query Match:

```
DB:                23          Gaps:          0
US-09-696-872-37 (1-4) x ABC27131 (1-13)
QY                1  LysAspGluLeu 4
DB                12  AAGGACGAGTTA 1
RESULT 6
ABC27132
ID  ABC27132 standard; DNA; 13 BP.
AC  ABC27132;
XX
XX
XX  20-FEB-2002 (first entry)
DT
DE  Oligonucleotide SEQ ID NO 27149 for detecting SNP TSC0007386.
XX
XX  SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
KW  peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
KW  central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
OS  Homo sapiens.
XX  WO200177384-A2.
XX  18-OCT-2001.
XX  PD
XX  PF  06-APR-2001; 2001WO-IB00713.
XX  PR  07-APR-2000; 2000DE-1019173.
XX  PA  (EPIG-) EPIGENOMICS AG.
XX  PI  Olek A, Piepenbrock C, Berlin K;
XX  DR  WPI; 2001-657177/75.
XX  PT  Set of oligonucleotides, useful for diagnosis and cell typing, is
XX  designed to detect single nucleotide polymorphisms and cytosine
XX  methylation status -
XX
XX  Claim 1; SEQ ID 27149; 29pp + Sequence Listing; German.
XX
XX  This invention describes novel oligonucleotide primers or peptide nucleic
XX  acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
XX  and cytosine methylation status in chemically pretreated genomic DNA. The
XX  oligonucleotides are used for diagnosis and/or prognosis of cancer and a
XX  range of diseases including immune system, gastrointestinal, respiratory,
XX  central nervous system, cardiovascular and metabolic disorders. The
XX  oligomers are also used for detecting cell type differentiation.
XX  CC  AB100010-AB182073 represent the oligomers described in the invention.
XX  CC  NOTE: The sequence data for this patent did not form part of the printed
XX  specification, but was obtained in electronic format from WIPO at
XX  ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 13 BP; 6 A; 1 C; 4 G; 2 T; 0 other;
SQ
Alignment Scores:
Pred. No.:      151      Length:      13
Score:          20.00     Matches:      4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    23      Indels:      0
DB:              Gaps:      0
US-09-696-872-37 (1-4) x ABC27132 (1-13)
QY                1  LysAspGluLeu 4
DB                2  AAGGACGAGTTA 13
QY                1  LysAspGluLeu 4
DB                2  AAGGACGAGTTA 13
```

```
RESULT 7
ABC27133/C
ID  ABC27133 standard; DNA; 13 BP.
XX
XX  ABC27133;
XX
XX  20-FEB-2002 (first entry)
DT
DE  Oligonucleotide SEQ ID NO 27150 for detecting SNP TSC0007386.
XX
XX  SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;
KW  peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;
KW  central nervous system; gastrointestinal; respiratory; immune; metabolic.
XX
OS  Homo sapiens.
XX  WO200177384-A2.
XX  18-OCT-2001.
XX  PD
XX  PF  06-APR-2001; 2001WO-IB00713.
XX  PR  07-APR-2000; 2000DE-1019173.
XX  PA  (EPIG-) EPIGENOMICS AG.
XX  PI  Olek A, Piepenbrock C, Berlin K;
XX  DR  WPI; 2001-657177/75.
XX  PT  Set of oligonucleotides, useful for diagnosis and cell typing, is
XX  designed to detect single nucleotide polymorphisms and cytosine
XX  methylation status -
XX
XX  Claim 1; SEQ ID 27150; 29pp + Sequence Listing; German.
XX
XX  This invention describes novel oligonucleotide primers or peptide nucleic
XX  acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)
XX  and cytosine methylation status in chemically pretreated genomic DNA. The
XX  oligonucleotides are used for diagnosis and/or prognosis of cancer and a
XX  range of diseases including immune system, gastrointestinal, respiratory,
XX  central nervous system, cardiovascular and metabolic disorders. The
XX  oligomers are also used for detecting cell type differentiation.
XX  CC  AB100010-AB182073 represent the oligomers described in the invention.
XX  CC  NOTE: The sequence data for this patent did not form part of the printed
XX  specification, but was obtained in electronic format from WIPO at
XX  ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 13 BP; 2 A; 4 C; 1 G; 6 T; 0 other;
SQ
Alignment Scores:
Pred. No.:      151      Length:      13
Score:          20.00     Matches:      4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    23      Indels:      0
DB:              Gaps:      0
US-09-696-872-37 (1-4) x ABC27133 (1-13)
QY                1  LysAspGluLeu 4
DB                12  AAGGACGAGTTA 1
RESULT 8
ABF76114
ID  ABF76114 standard; DNA; 13 BP.
XX
XX  ABF76114;
XX
XX  22-FEB-2002 (first entry)
DT
XX
```

DE Oligonucleotide SEQ ID NO 176111 for detecting SNP TSC0043719.  
XX  
XX SNP: single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;  
KM peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;  
KW central nervous system; gastrointestinal; respiratory; immune; metabolic.  
XX  
OS Homo sapiens.  
XX  
XX WO200177384-A2.  
XX  
XX 18-OCT-2001.  
XX  
XX 06-APR-2001; 2001WO-IB00713.  
XX  
XX 07-APR-2000; 2000DE-1019173.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2001-657177/75.  
XX  
XX Set of oligonucleotides, useful for diagnosis and cell typing, is  
PT designed to detect single nucleotide polymorphisms and cytosine  
PT methylation status -  
XX  
XX Claim 1; SEQ ID 176111; 29pp + Sequence Listing; German.  
XX  
XX This invention describes novel oligonucleotide primers or peptide nucleic  
CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)  
CC and cytosine methylation status in chemically pretreated genomic DNA. The  
CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a  
CC range of diseases including immune system, gastrointestinal, respiratory,  
CC central nervous system, cardiovascular and metabolic disorders. The  
CC oligomers are also used for detecting cell type differentiation.  
CC ABC00010-ABC99989, ABP00010-ABP99989, ABH00010-ABH99989 and  
CC AB100010-AB182073 represent the oligomers described in the invention.  
CC NOTE: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 13 BP; 5 A; 0 C; 5 G; 3 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 151 Length: 13  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 23 Gaps: 0  
US-09-696-872-37 (1-4) x ABF76114 (1-13)  
QY 1 LysApGluLeu 4  
DB 1 AAAGATGAGTTG 12  
RESULT 9  
ID ABF76115 standard; DNA; 13 BP.  
XX AC ABF76115;  
XX 22-FEB-2002 (first entry)  
XX  
XX Oligonucleotide SEQ ID NO 176112 for detecting SNP TSC0043719.  
XX  
XX SNP: single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;  
KW peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;  
KW central nervous system; gastrointestinal; respiratory; immune; metabolic.  
XX  
XX Homo sapiens.  
OS

PN WO200177384-A2.  
XX  
XX 18-OCT-2001.  
XX  
XX 06-APR-2001; 2001WO-IB00713.  
XX  
XX 07-APR-2000; 2000DE-1019173.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2001-657177/75.  
XX  
XX Set of oligonucleotides, useful for diagnosis and cell typing, is  
PT designed to detect single nucleotide polymorphisms and cytosine  
PT methylation status -  
XX  
XX Claim 1; SEQ ID 176112; 29pp + Sequence Listing; German.  
XX  
XX This invention describes novel oligonucleotide primers or peptide nucleic  
CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)  
CC and cytosine methylation status in chemically pretreated genomic DNA. The  
CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a  
CC range of diseases including immune system, gastrointestinal, respiratory,  
CC central nervous system, cardiovascular and metabolic disorders. The  
CC oligomers are also used for detecting cell type differentiation.  
CC ABC00010-ABC99989, ABP00010-ABP99989, ABH00010-ABH99989 and  
CC AB100010-AB182073 represent the oligomers described in the invention.  
CC NOTE: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 13 BP; 3 A; 5 C; 0 G; 5 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 151 Length: 13  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 23 Gaps: 0  
US-09-696-872-37 (1-4) x ABF76115 (1-13)  
QY 1 LysApGluLeu 4  
DB 13 AAAGATGAGTTG 2  
RESULT 10  
ID ABF76118 standard; DNA; 13 BP.  
XX AC ABF76118;  
XX 22-FEB-2002 (first entry)  
XX  
XX Oligonucleotide SEQ ID NO 176115 for detecting SNP TSC0043719.  
XX  
XX SNP: single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;  
KW peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;  
KW central nervous system; gastrointestinal; respiratory; immune; metabolic.  
XX  
XX Homo sapiens.  
XX  
XX WO200177384-A2.  
XX  
XX 18-OCT-2001.  
XX  
XX 06-APR-2001; 2001WO-IB00713.  
XX  
XX 07-APR-2000; 2000DE-1019173.  
XX

PA (EPiG-) EPIGENOMICS AG.  
 XX Olek A, Piepenbrock C, Berlin K;  
 XX WPI; 2001-657177/75.  
 DR Set of oligonucleotides, useful for diagnosis and cell typing, is  
 XX designed to detect single nucleotide polymorphisms and cytosine  
 XX methylation status -  
 XX  
 PS Claim 1; SEQ ID 176115; 29pp + Sequence Listing; German.  
 CC This invention describes novel oligonucleotide primers or peptide nucleic  
 CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)  
 CC and cytosine methylation status in chemically pretreated genomic DNA. The  
 CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a  
 CC range of diseases including immune system, gastrointestinal, respiratory,  
 CC central nervous system, cardiovascular and metabolic disorders. The  
 CC oligomers are also used for detecting cell type differentiation.  
 CC ABC00010-ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and  
 CC AB100010-AB182073 represent the oligomers described in the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 13 BP; 5 A; 1 C; 5 G; 2 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 151 Length: 13  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0  
 US-09-696-872-37 (1-4) x ABF76118 (1-13)  
 QY 1 LysAaPGLuLeu 4  
 DB 1 AAAGACGAGTTG 12  
 RESULT 11  
 ABF76119/C  
 ID ABF76119 standard; DNA; 13 BP.  
 AC ABF76119;  
 XX  
 DT 22-FEB-2002 (first entry)  
 DE Oligonucleotide SEQ ID NO 176116 for detecting SNP TSC0043719.  
 XX  
 KW SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;  
 KW peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;  
 KW central nervous system; gastrointestinal; respiratory; immune; metabolic.  
 OS Homo sapiens.  
 XX  
 PN WO200177384-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 06-APR-2001; 2001WO-IB00713.  
 XX  
 PR 07-APR-2000; 2000DE-1019173.  
 XX  
 PA (EPiG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2001-657177/75.  
 XX  
 CC Set of oligonucleotides, useful for diagnosis and cell typing, is  
 CC designed to detect single nucleotide polymorphisms and cytosine

PT methylation status -  
 XX  
 PS Claim 1; SEQ ID 176116; 29pp + Sequence Listing; German.  
 XX  
 CC This invention describes novel oligonucleotide primers or peptide nucleic  
 CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)  
 CC and cytosine methylation status in chemically pretreated genomic DNA. The  
 CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a  
 CC range of diseases including immune system, gastrointestinal, respiratory,  
 CC central nervous system, cardiovascular and metabolic disorders. The  
 CC oligomers are also used for detecting cell type differentiation.  
 CC ABC00010-ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and  
 CC AB100010-AB182073 represent the oligomers described in the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 13 BP; 2 A; 5 C; 1 G; 5 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 151 Length: 13  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0  
 US-09-696-872-37 (1-4) x ABF76119 (1-13)  
 QY 1 LysAaPGLuLeu 4  
 DB 13 AAAGACGAGTTG 2  
 RESULT 12  
 ABH29150  
 ID ABH29150 standard; DNA; 13 BP.  
 AC ABH29150;  
 XX  
 DT 22-FEB-2002 (first entry)  
 DE Oligonucleotide SEQ ID NO 229127 for detecting SNP TSC0055897.  
 XX  
 KW SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;  
 KW peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;  
 KW central nervous system; gastrointestinal; respiratory; immune; metabolic.  
 OS Homo sapiens.  
 XX  
 PN WO200177384-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 06-APR-2001; 2001WO-IB00713.  
 XX  
 PR 07-APR-2000; 2000DE-1019173.  
 XX  
 PA (EPiG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2001-657177/75.  
 XX  
 CC Set of oligonucleotides, useful for diagnosis and cell typing, is  
 CC designed to detect single nucleotide polymorphisms and cytosine  
 CC methylation status -  
 CC  
 PS Claim 1; SEQ ID 229127; 29pp + Sequence Listing; German.  
 XX  
 CC This invention describes novel oligonucleotide primers or peptide nucleic  
 CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)  
 CC and cytosine methylation status in chemically pretreated genomic DNA. The  
 CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a



CC range of diseases including immune system, gastrointestinal, respiratory,  
 CC central nervous system, cardiovascular and metabolic disorders. The  
 CC oligomers are also used for detecting cell type differentiation.  
 CC AB000010-ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and  
 CC AB100010-AB12073 represent the oligomers described in the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 13 BP; 7 A; 0 C; 3 G; 3 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 151 Length: 13  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-09-696-872-37 (1-4) x ABH29150 (1-13)  
 OY 1 LysAspGluLeu 4  
 DB 2 AAAGATGATTG 13  
 RESULT 13  
 ABH29151/c  
 ID ABH29151 standard; DNA; 13 BP.  
 XX  
 AC ABH29151;  
 XX  
 DT 22-FEB-2002 (first entry)  
 XX  
 DE Oligonucleotide SEQ ID NO 229128 for detecting SNP TSC0055897.  
 XX  
 KM SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;  
 KM peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;  
 KM central nervous system; gastrointestinal; respiratory; immune; metabolic.  
 OS Homo sapiens.  
 XX  
 PN WO200177384-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 06-APR-2001; 2001WO-IB00713.  
 XX  
 PR 07-APR-2000; 2000DE-1019173.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 PS WPI; 2001-657177/75.  
 XX  
 DR Set of oligonucleotides, useful for diagnosis and cell typing, is  
 PT designed to detect single nucleotide polymorphisms and cytosine  
 PT methylation status -  
 XX  
 XX Claim 1; SEQ ID 229128; 29pp + Sequence Listing; German.  
 CC This invention describes novel oligonucleotide primers or peptide nucleic  
 CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)  
 CC and cytosine methylation status in chemically pretreated genomic DNA. The  
 CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a  
 CC range of diseases including immune system, gastrointestinal, respiratory,  
 CC central nervous system, cardiovascular and metabolic disorders. The  
 CC oligomers are also used for detecting cell type differentiation.  
 CC AB000010-ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and  
 CC AB100010-AB12073 represent the oligomers described in the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 13 BP; 3 A; 3 C; 0 G; 7 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 151 Length: 13  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-09-696-872-37 (1-4) x ABH29151 (1-13)  
 OY 1 LysAspGluLeu 4  
 DB 12 AAAGATGATTG 1  
 RESULT 14  
 ABH29152  
 ID ABH29152 standard; DNA; 13 BP.  
 XX  
 AC ABH29152;  
 XX  
 DT 22-FEB-2002 (first entry)  
 XX  
 DE Oligonucleotide SEQ ID NO 229129 for detecting SNP TSC0055897.  
 XX  
 KM SNP; single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;  
 KM peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;  
 KM central nervous system; gastrointestinal; respiratory; immune; metabolic.  
 OS Homo sapiens.  
 XX  
 PN WO200177384-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 06-APR-2001; 2001WO-IB00713.  
 XX  
 PR 07-APR-2000; 2000DE-1019173.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 PS WPI; 2001-657177/75.  
 XX  
 DR Set of oligonucleotides, useful for diagnosis and cell typing, is  
 PT designed to detect single nucleotide polymorphisms and cytosine  
 PT methylation status -  
 XX  
 XX Claim 1; SEQ ID 229129; 29pp + Sequence Listing; German.  
 CC This invention describes novel oligonucleotide primers or peptide nucleic  
 CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)  
 CC and cytosine methylation status in chemically pretreated genomic DNA. The  
 CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a  
 CC range of diseases including immune system, gastrointestinal, respiratory,  
 CC central nervous system, cardiovascular and metabolic disorders. The  
 CC oligomers are also used for detecting cell type differentiation.  
 CC AB000010-ABC99989, ABF00010-ABF99989, ABH00010-ABH99989 and  
 CC AB100010-AB12073 represent the oligomers described in the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 13 BP; 7 A; 1 C; 3 G; 2 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 151 Length: 13  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0  
 DB: 23 Gaps: 0  
 US-09-696-872-37 (1-4) x ABH29152 (1-13)

QY 1 LysApGjuleu 4  
 |||||  
 Db 2 AAAGACGAATTG 13

## RESULT 15

ABH29153/C  
 ID ABH29153 standard; DNA, 13 BP.

AC ABH29153;  
 XX

DT 22-FEB-2002 (first entry)  
 XX

DE Oligonucleotide SEQ ID NO 229130 for detecting SNP TSC055897.

XX SNP, single nucleotide polymorphism; human; diagnosis; PNA; cancer; CNS;  
 KM peptide nucleic acid; cytosine methylation; cardiovascular; primer; ss;  
 KW central nervous system; gastrointestinal; respiratory; immune; metabolic.  
 XX

OS Homo sapiens.  
 XX

PN WO200177384-A2.  
 XX

PD 18-OCT-2001.  
 XX

PF 06-APR-2001; 2001WO-1B00713.  
 XX

PR 07-APR-2000; 2000DE-1019173.  
 XX

PA (EPIC-) EPIGENOMICS AG.  
 XX

PI Olek A, Piepenbrock C, Berlin K;  
 XX

DR WPI; 2001-657177/75.  
 XX

PT Set of oligonucleotides, useful for diagnosis and cell typing, is  
 PT designed to detect single nucleotide polymorphisms and cytosine  
 PT methylation status -

PS Claim 1; SEQ ID 229130; 29pp + Sequence Listing; German.  
 XX

CC This invention describes novel oligonucleotide primers or peptide nucleic  
 CC acid (PNA) oligomers for detecting single nucleotide polymorphisms (SNP)  
 CC and cytosine methylation status in chemically pretreated genomic DNA. The  
 CC oligonucleotides are used for diagnosis and/or prognosis of cancer and a  
 CC range of diseases including immune system, gastrointestinal, respiratory,  
 CC central nervous system, cardiovascular and metabolic disorders. The  
 CC oligomers are also used for detecting cell type differentiation.

CC AB100010-AB182073 represent the oligomers described in the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 13 BP; 2 A; 3 C; 1 G; 7 T; 0 other;

## Alignment Scores:

Pred. No.:	151	Length:	13
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	23	Gaps:	0

US-09-696-872-37 (1-4) x ABH29153 (1-13)

QY 1 LysApGjuleu 4  
 |||||  
 Db 12 AAAGACGAATTG 1

Search completed: August 26, 2003, 17:06:38  
 Job time : 14.6991 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 26, 2003, 16:55:35 ; Search time 3.15044 Seconds

(without alignments)  
560.408 Million cell updates/sec

Title: US-09-696-872-37

Perfect score: 20

Sequence: 1 KDEL 4

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Xgapop 10.0	10.0	10.0	0.5
Fgapop 6.0	6.0	6.0	7.0
Delop 6.0	6.0	6.0	7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-O=/cgn2\_1/USPFO\_spool/US09696872/runat\_26082003\_151138\_3255/app\_query.fasta\_1.462  
-DB=Issued\_Patents\_NA -QEMT=fastap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Dlosum62 -TRANS=human40.cdt  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORML=ext -HEAPSIZE=500 -MITLEN=0 -MAXLEN=2000000000  
-USER=US09696872@cgn2\_1.1.76@runat\_26082003\_151138\_3255 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELTEXT=7

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	12	4	US-08-983-564A-24
2	20	100.0	16	4	US-09-371-772B-7062
3	20	100.0	18	4	US-09-253-911-38
4	20	100.0	20	3	US-09-280-805-229
5	20	100.0	20	3	US-09-280-805-230
6	20	100.0	20	3	US-08-875-847B-16
7	20	100.0	20	4	US-09-378-842-16
8	20	100.0	20	4	US-09-702-327-66
9	20	100.0	20	4	US-09-702-327-67
10	20	100.0	20	4	US-09-702-327-68
11	20	100.0	21	1	US-08-052-997-6
12	20	100.0	21	1	US-08-684-672-6

13	20	100.0	22	3	US-09-191-852-5	Sequence 5, Appli
14	20	100.0	22	4	US-07-974-409C-315	Sequence 315, App
15	20	100.0	22	4	US-07-974-409C-316	Sequence 316, App
16	20	100.0	22	4	US-07-974-409C-317	Sequence 317, App
17	20	100.0	22	4	US-08-817-906-5	Sequence 5, Appli
18	20	100.0	22	5	PCT-US93-00977-315	Sequence 315, App
19	20	100.0	22	5	PCT-US93-00977-316	Sequence 316, App
20	20	100.0	22	5	PCT-US93-00977-317	Sequence 317, App
21	20	100.0	22	5	PCT-US95-13376-5	Sequence 5, Appli
22	20	100.0	24	4	US-09-471-573A-10	Sequence 10, Appl
23	20	100.0	24	4	US-09-471-573A-20	Sequence 20, Appl
24	20	100.0	25	3	US-09-050-159-109	Sequence 109, App
25	20	100.0	26	3	US-09-191-852-6	Sequence 6, Appli
26	20	100.0	26	5	PCT-US95-13376-6	Sequence 6, Appli
27	20	100.0	27	4	US-08-198-603C-18	Sequence 18, Appl
28	20	100.0	30	4	US-09-471-573A-9	Sequence 9, Appli
29	20	100.0	31	4	US-08-882-164D-42	Sequence 42, Appl
30	20	100.0	31	4	US-08-679-645-194	Sequence 394, App
31	20	100.0	32	3	US-09-243-810-4	Sequence 4, Appli
32	20	100.0	32	3	US-09-191-852-7	Sequence 7, Appli
33	20	100.0	38	4	US-08-817-906-7	Sequence 7, Appli
34	20	100.0	38	5	PCT-US95-13376-7	Sequence 7, Appli
35	20	100.0	41	4	US-08-840-713-7	Sequence 7, Appli
36	20	100.0	42	3	US-08-947-965-56	Sequence 56, App
37	20	100.0	43	1	US-07-931-473B-256	Sequence 256, App
38	20	100.0	43	1	US-07-714-131C-256	Sequence 256, App
39	20	100.0	43	1	US-08-412-110-256	Sequence 256, App
40	20	100.0	43	1	US-08-409-442A-256	Sequence 256, App
41	20	100.0	43	1	US-08-469-609A-256	Sequence 256, App
42	20	100.0	43	3	US-09-143-190-256	Sequence 256, App
43	20	100.0	43	4	US-09-502-344-256	Sequence 256, App
44	20	100.0	43	4	US-08-840-713-8	Sequence 8, Appli
45	20	100.0				

#### ALIGNMENTS

RESULT 1  
US-08-983-564A-24  
; Sequence 24, Application US/08983564A  
; Patent No. 6344600  
; GENERAL INFORMATION:  
; APPLICANT: Meroc, Bertrand  
; APPLICANT: Dieryck, Wilfrid  
; APPLICANT: Lenee, Philippe  
; APPLICANT: Warden, Michel  
; APPLICANT: Gaudier, Veronique  
; APPLICANT: Pagnier, Renee-Jossee  
; APPLICANT: Baudino, Sylvie  
; APPLICANT: Poyart, Claude  
; TITLE OF INVENTION: METHOD FOR PRODUCING HAEMIN PROTEINS USING PLANT CELLS,  
; TITLE OF INVENTION: RESULTING PROTEINS AND PRODUCTS CONTAINING SAME  
; FILE REFERENCE: 8076.147USWO  
; CURRENT APPLICATION NUMBER: US/08/983, 564A  
; CURRENT FILING DATE: 1998-06-09  
; PRIOR APPLICATION NUMBER: PCT/FR96/01123  
; PRIOR FILING DATE: 1996-07-17  
; PRIOR APPLICATION NUMBER: 95/08615  
; PRIOR FILING DATE: 1995-07-17  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 12  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Homo sapiens  
US-08-983-564A-24  
Alignment Scores: 32  
Pred. No.: 20.00  
Score: 20.00  
Length: 12  
Matches: 4  
Percent Similarity: 100.00%  
Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0  
US-09-696-872-37 (1-4) x US-08-983-564A-24 (1-12)

QY 1 LysaApGjuleu 4  
DB 1 AAAGATGAGCTA 12

RESULT 2  
US-09-371-772B-7062  
; Sequence 7062, Application US/09371772B  
; Patent No. 6566127  
; GENERAL INFORMATION:  
; APPLICANT: Ribozyme Pharmaceuticals, Inc.  
; APPLICANT: Pavco, Pam  
; APPLICANT: McSwigen, Jim  
; APPLICANT: Stinchcomb, Dan  
; APPLICANT: Escobedo, Jaime  
; TITLE OF INVENTION: Method and Reagent for the Treatment of Diseases or Conditions Re  
; FILE REFERENCE: MBH00, 876-J (237/198)  
; CURRENT APPLICATION NUMBER: US/09/371, 772B  
; CURRENT FILING DATE: 1999-08-10  
; PRIOR APPLICATION NUMBER: US 60/005, 974  
; PRIOR FILING DATE: 1995-10-26  
; PRIOR APPLICATION NUMBER: US 08/584, 040  
; PRIOR FILING DATE: 1996-01-08  
; NUMBER OF SEQ ID NOS: 14225  
; SOFTWARE: Patent version 3.0  
; SEQ ID NO 7062  
; LENGTH: 16  
; TYPE: RNA  
; ORGANISM: Homo sapiens  
US-09-371-772B-7062

Alignment Scores:  
Pred. No.: 44.1 Length: 16  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0  
US-09-696-872-37 (1-4) x US-09-371-772B-7062 (1-16)

QY 1 LysaApGjuleu 4  
DB 5 AAAGACGAACTU 16

RESULT 3  
US-09-255-911-38  
; Sequence 38, Application US/09255911  
; Patent No. 6013522  
; GENERAL INFORMATION:  
; APPLICANT: Brett P. Monia  
; APPLICANT: Lex M. Cowart  
; TITLE OF INVENTION: ANTISENSE MODULATION OF SMAD1 EXPRESSION  
; FILE REFERENCE: RTS-0040  
; CURRENT APPLICATION NUMBER: US/09/255, 911  
; CURRENT FILING DATE: 1999-02-23  
; NUMBER OF SEQ ID NOS: 46  
; SEQ ID NO 38  
; LENGTH: 18  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Antisense Oligonucleotide  
US-09-255-911-38

Alignment Scores:  
Pred. No.: 50.3 Length: 18

Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0  
US-09-696-872-37 (1-4) x US-09-255-911-38 (1-18)

QY 1 LysaApGjuleu 4  
DB 1 AAAGATGAGCTC 12

RESULT 4  
US-09-280-805-229  
; Sequence 229, Application US/09280805  
; Patent No. 6184212  
; GENERAL INFORMATION:  
; APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.  
; APPLICANT: Graham, Brett P. Monia  
; TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN MDM2  
; NUMBER OF SEQUENCES: 271  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Law Offices of Jane Massey Licata  
; STREET: 66 East Main Street  
; CITY: Marlton  
; STATE: NJ  
; COUNTRY: U.S.A.  
; ZIP: 08053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
; COMPUTER: IBM PC  
; OPERATING SYSTEM: WINDOWS 95  
; SOFTWARE: WORDPERFECT 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/280, 805  
; FILING DATE: herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/048, 810  
; FILING DATE: March 26, 1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Licata, Jane Massey  
; REGISTRATION NUMBER: 32,257  
; REFERENCE/DOCKET NUMBER: ISPH-0346  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 609-810-1515  
; TELEFAX: 609-810-1454  
; INFORMATION FOR SEQ ID NO: 229:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: Linear  
; ANTI-SENSE: Yes  
US-09-280-805-229

Alignment Scores:  
Pred. No.: 56.6 Length: 20  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0  
US-09-696-872-37 (1-4) x US-09-280-805-229 (1-20)

QY 1 LysaApGjuleu 4  
DB 2 AAAGATGAGCTA 13

RESULT 5  
US-09-280-805-230

Sequence 230, Application US/09280805  
Patent No. 6184212  
GENERAL INFORMATION:  
APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.  
APPLICANT: Graham, Brett P. Monia  
TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN MDM2  
TITLE OF INVENTION: EXPRESSION  
NUMBER OF SEQUENCES: 271  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Jane Massey Licata  
STREET: 66 East Main Street  
CITY: Marlton  
STATE: NJ  
COUNTRY: U.S.A.  
ZIP: 08053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM PC  
OPERATING SYSTEM: WINDOWS 95  
SOFTWARE: WORDPERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/280,805  
FILING DATE: herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/048,810  
FILING DATE: March 26, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Licata, Jane Massey  
REGISTRATION NUMBER: 32,257  
REFERENCE/DOCKET NUMBER: ISPH-0346  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 609-810-1515  
TELEFAX: 609-810-1454  
INFORMATION FOR SEQ ID NO: 230:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
ANTI-SENSE: Yes  
US-09-280-805-230

Alignment Scores:  
Pred. No.: 56.6 Length: 20  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-09-696-872-37 (1-4) x US-09-280-805-230 (1-20)

QY 1 LysAspGluLeu 4  
Db 6 AAAGATGAGCTA 17

RESULT 6  
US-08-875-847B-16  
Sequence 16, Application US/08875847B  
Patent No. 6255105  
GENERAL INFORMATION:  
APPLICANT: The Government of the United  
APPLICANT: States of America as represented by the  
APPLICANT: Secretary, Department of Health and Human  
APPLICANT: Services; Callahan, Robert; Marchetti,  
APPLICANT: Antonio; Buttitta, Fiamma; Smith, Gilbert H.  
TITLE OF INVENTION: Nucleotide And Deduced  
TITLE OF INVENTION: Amino Acid Sequences Of A New Tumor Gene,  
TITLE OF INVENTION: Int6, And the Use Of Reagents Derived From  
TITLE OF INVENTION: These Sequences In Diagnostic Assays,  
TITLE OF INVENTION: Vaccines, Immunotherapy And Gene Therapy  
NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS WORD 97  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/875,847B  
FILING DATE: 09-FEB-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/385,998  
FILING DATE: 09-FEB-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: William S. Feiler  
REGISTRATION NUMBER: 26,728  
REFERENCE/DOCKET NUMBER: 2026-4179PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-875-847B-16

Alignment Scores:  
Pred. No.: 56.6 Length: 20  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-09-696-872-37 (1-4) x US-08-875-847B-16 (1-20)

QY 1 LysAspGluLeu 4  
Db 7 AAAGACGACTC 18

RESULT 7  
US-09-378-842-16  
Sequence 16, Application US/09378842  
Patent No. 6342392  
GENERAL INFORMATION:  
APPLICANT: The Government of the United  
APPLICANT: States of America as represented by the  
APPLICANT: Secretary, Department of Health and Human  
APPLICANT: Services; Callahan, Robert; Marchetti,  
APPLICANT: Antonio; Buttitta, Fiamma; Smith, Gilbert H.  
TITLE OF INVENTION: Nucleotide And Deduced  
TITLE OF INVENTION: Amino Acid Sequences Of A New Tumor Gene,  
TITLE OF INVENTION: Int6, And the Use Of Reagents Derived From  
TITLE OF INVENTION: These Sequences In Diagnostic Assays,  
TITLE OF INVENTION: Vaccines, Immunotherapy And Gene Therapy  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154



RESULT 11  
US-08-052-997-6  
; Sequence 6, Application US/08052997  
; Patent No. 5556786  
; GENERAL INFORMATION:  
; APPLICANT: Kere, Juba  
; APPLICANT: Schlessinger, David  
; APPLICANT: de la Chapelle, Albert  
; TITLE OF INVENTION: ANHIDROTIC ECTODERMAL DYSPLASIA GENE  
; TITLE OF INVENTION: AND METHOD OF DETECTING SAME  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: POPHAM HAIR SCHNOBIRCH & KAUFMAN, LTD.  
; STREET: 1225 Eye Street N.W., Suite 1000  
; CITY: Washington, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/052,997  
; FILING DATE: 27-APR-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: O'Shaughnessy, Brian P.  
; REGISTRATION NUMBER: 32,747  
; REFERENCE/DOCKET NUMBER: 9594/81-2189  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 289-1200  
; TELEFAX: (202) 289-6674  
; TELEX: 248516  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHEICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: Yeast artificial chromosome  
; US-08-052-997-6  
; Alignment Scores:  
; Pred. No.: 59.8 Length: 21  
; Score: 20.00 Matches: 4  
; Percent Similarity: 100.00% Conservative: 0  
; Best Local Similarity: 100.00% Mismatches: 0  
; Query Match: 100.00% Indels: 0  
; DB: 1 Gaps: 0  
US-09-696-872-37 (1-4) x US-08-052-997-6 (1-21)  
Qy 1 Lysaapgiuleu 4  
Db 8 AAAGATGAGCTG 19  
RESULT 12  
US-08-684-672-6  
; Sequence 6, Application US/08684672  
; Patent No. 5700926  
; GENERAL INFORMATION:  
; APPLICANT: KERE, Juba  
; APPLICANT: SCHLESSINGER, David  
; APPLICANT: de la CHAPELLE, Albert  
; APPLICANT: SRIVASTAVA, Anand Kumar  
; TITLE OF INVENTION: MOLECULAR CLONING OF THE ANHIDROTIC  
; TITLE OF INVENTION: ECTODERMAL DYSPLASIA GENE

NUMBER OF SEQUENCES: 36  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SNECKER & MATHIS, L.L.P.  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 2213-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/684,672  
FILING DATE: 22-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/052,997  
FILING DATE: 27-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Shaughnessy, Brian P.  
REGISTRATION NUMBER: 32,747  
REFERENCE/DOCKET NUMBER: 030956-002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-684-672-6  
; Alignment Scores:  
; Pred. No.: 59.8 Length: 21  
; Score: 20.00 Matches: 4  
; Percent Similarity: 100.00% Conservative: 0  
; Best Local Similarity: 100.00% Mismatches: 0  
; Query Match: 100.00% Indels: 0  
; DB: 1 Gaps: 0  
US-09-696-872-37 (1-4) x US-08-684-672-6 (1-21)  
Qy 1 Lysaapgiuleu 4  
Db 8 AAAGATGAGCTG 19  
RESULT 13  
US-09-191-852-5  
; Sequence 5, Application US/09191852  
; Patent No. 6194560  
; GENERAL INFORMATION:  
; APPLICANT: Charles J. Arntzen, Hugh S. Mason, and Tariq A. Haq  
; TITLE OF INVENTION: ORAL IMMUNIZATION WITH TRANSGENIC PLANTS  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski L.L.P.  
; STREET: 1301 McKinney, Suite 5100  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77010  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/191,852  
; FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13376  
FILING DATE: 24-OCT-1995  
APPLICATION NUMBER: 08/817,906  
FILING DATE: 04-AUG-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Fox, David L.  
REGISTRATION NUMBER: 40,612  
REFERENCE/DOCKET NUMBER: P01590US1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 713-651-5151  
TELEFAX: 713-651-5246  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-191-852-5

Alignment Scores:  
Pred. No.: 63 Length: 22  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 3 Gaps: 0

US-09-696-872-37 (1-4) x US-09-191-852-5 (1-22)

QY 1 LysaepgJlueu 4  
DB 8 AAAGATGAGCTA 19

RESULT 14  
US-07-974-409C-315/c  
Sequence 315, Application US/07974409C  
Patent No. 6300058  
GENERAL INFORMATION:  
APPLICANT: Akitaya, Tatsuo  
APPLICANT: Mitsuhashi, Masato  
APPLICANT: Cooper, Allan  
TITLE OF INVENTION: METHOD AND REAGENT  
TITLE OF INVENTION: FOR MEASURING MESSENGER RNA  
NUMBER OF SEQUENCES: 457  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson, and Bear  
STREET: 620 Newport Center Dr. Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/974,409C  
FILING DATE: 12-NOV-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E.  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: HITACHI.006CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
INFORMATION FOR SEQ ID NO: 315:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22  
TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-07-974-409C-315

Alignment Scores:  
Pred. No.: 63 Length: 22  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0

US-09-696-872-37 (1-4) x US-07-974-409C-315 (1-22)

QY 1 LysaepgJlueu 4  
DB 15 AAAGACGAACCTC 4

RESULT 15  
US-07-974-409C-316/c  
Sequence 316, Application US/07974409C  
Patent No. 6300058  
GENERAL INFORMATION:  
APPLICANT: Akitaya, Tatsuo  
APPLICANT: Mitsuhashi, Masato  
APPLICANT: Cooper, Allan  
TITLE OF INVENTION: METHOD AND REAGENT  
TITLE OF INVENTION: FOR MEASURING MESSENGER RNA  
NUMBER OF SEQUENCES: 457  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson, and Bear  
STREET: 620 Newport Center Dr. Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: USA  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/974,409C  
FILING DATE: 12-NOV-1992  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Altman, Daniel E.  
REGISTRATION NUMBER: 34,115  
REFERENCE/DOCKET NUMBER: HITACHI.006CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 714-760-0404  
TELEFAX: 714-760-9502  
INFORMATION FOR SEQ ID NO: 316:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-07-974-409C-316

Alignment Scores:  
Pred. No.: 63 Length: 22  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 4 Gaps: 0



US-09-696-872-37 (1-4) x US-07-974-408C-316 (1-22)

OY 1 LysApGluLeu 4

Db 15 AAGAGCGAACTC 4

Search completed: August 26, 2003, 19:22:50  
Job time : 4.15044 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 26, 2003, 18:30:26 ; Search time 11.5044 Seconds  
(without alignments)  
781.669 Million cell updates/sec

Title: US-09-696-872-37  
Perfect score: 20  
Sequence: 1 KDELL 4

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1517243 seqs, 1124081882 residues  
Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09686872/runat\_26082003\_151139\_3337/app\_query.fasta\_1.462  
-DB=Published Applications\_NA -QFMT=faetap -SUFFIX=trmpb -MINMATCH=0.1  
-LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=dl0sum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MOBE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09686872.@CGN1.1.291@runat\_26082003\_151139\_3337  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERT -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONELOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PC1\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PC1US\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	12	9	US-09-333-527-1 Sequence 1, Appl

2	20	100.0	12	13	US-10-085-853-24	Sequence 24, Appl
3	20	100.0	12	14	US-10-127-427-1	Sequence 1, Appl
4	20	100.0	15	14	US-10-287-919-1514	Sequence 1514, Ap
5	20	100.0	15	14	US-10-287-919-1515	Sequence 1515, Ap
6	20	100.0	15	14	US-10-287-919-2042	Sequence 2042, Ap
7	20	100.0	15	14	US-10-287-919-2043	Sequence 2043, Ap
8	20	100.0	17	11	US-09-780-533A-652	Sequence 652, App
9	20	100.0	17	11	US-09-780-533A-653	Sequence 653, App
10	20	100.0	17	11	US-09-780-533A-654	Sequence 654, App
11	20	100.0	17	11	US-09-780-533A-1501	Sequence 1501, Ap
12	20	100.0	19	10	US-09-969-373-3751	Sequence 3751, Ap
13	20	100.0	20	9	US-09-752-983-229	Sequence 229, App
14	20	100.0	20	9	US-09-752-983-230	Sequence 230, App
15	20	100.0	20	10	US-09-791-406-66	Sequence 66, Appl
16	20	100.0	20	10	US-09-791-406-67	Sequence 67, Appl
17	20	100.0	20	10	US-09-791-406-68	Sequence 68, Appl
18	20	100.0	20	11	US-09-858-152A-16	Sequence 16, Appl
19	20	100.0	22	10	US-09-969-373-3571	Sequence 3571, Ap
20	20	100.0	23	9	US-09-866-778A-10	Sequence 10, Appl
21	20	100.0	23	9	US-09-866-778A-11	Sequence 11, Appl
22	20	100.0	23	14	US-10-002-292A-7	Sequence 7, Appl
23	20	100.0	24	12	US-10-079-167-3	Sequence 3, Appl
24	20	100.0	24	12	US-10-286-628-37	Sequence 37, Appl
25	20	100.0	25	10	US-09-263-959-1077	Sequence 1077, Ap
26	20	100.0	25	14	US-10-215-112-2977	Sequence 2977, Ap
27	20	100.0	25	14	US-10-215-112-12022	Sequence 12022, A
28	20	100.0	25	14	US-10-098-263B-1698	Sequence 1698, Ap
29	20	100.0	25	14	US-10-098-263B-5107	Sequence 5107, Ap
30	20	100.0	25	14	US-10-098-263B-9595	Sequence 9595, Ap
31	20	100.0	25	14	US-10-098-263B-10231	Sequence 10231, A
32	20	100.0	25	14	US-10-098-263B-10867	Sequence 10867, A
33	20	100.0	25	14	US-10-098-263B-17233	Sequence 17233, A
34	20	100.0	25	14	US-10-098-263B-17234	Sequence 17234, A
35	20	100.0	25	14	US-10-098-263B-27417	Sequence 27417, A
36	20	100.0	25	14	US-10-098-263B-31915	Sequence 31915, A
37	20	100.0	25	14	US-10-098-263B-50278	Sequence 50278, A
38	20	100.0	25	14	US-10-098-263B-58369	Sequence 58369, A
39	20	100.0	25	14	US-10-098-263B-60887	Sequence 60887, A
40	20	100.0	25	14	US-10-098-263B-66097	Sequence 66097, A
41	20	100.0	25	14	US-10-098-263B-66666	Sequence 66666, A
42	20	100.0	25	14	US-10-098-263B-70671	Sequence 70671, A
43	20	100.0	25	14	US-10-098-263B-71415	Sequence 71415, A
44	20	100.0	25	14	US-10-098-263B-76429	Sequence 76429, A
45	20	100.0	25	14	US-10-098-263B-96476	Sequence 96476, A

#### ALIGNMENTS

RESULT 1  
US-09-333-527-1  
Sequence 1, Application US/09333527  
Patent No. US20020078472A1  
GENERAL INFORMATION:  
APPLICANT: Paul CHRISTOU; Eva STROGER; Rainer FISCHER; Carmen MARTIN-VAQUERO; Stef  
TITLE OF INVENTION: METHODS AND MEANS FOR EXPRESSION OF MAMMALIAN POLYPEPTIDES  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pubright & Jaworski L.L.P.  
STREET: 666 Fifth Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/333,527  
FILING DATE: Concurrently Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/089,322  
FILING DATE: June 15, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary Anne Schofield  
REGISTRATION NUMBER: 36,669  
REFERENCE/DOCKET NUMBER: KL/JIC 202.1 - JEL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 318-3000  
TELEFAX: (212) 752-5958  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-333-527-1

Alignment Scores:  
Pred. No.: 190  
Score: 20.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 9  
Gaps: 0

US-09-696-872-37 (1-4) x US-09-333-527-1 (1-12)

QY 1 LysAspGluLeu 4  
|||  
1 AAAGATGAGCTC 12

Db

RESULT 2  
US-10-085-853-24  
Sequence 24, Application US/10085853  
GENERAL INFORMATION:  
APPLICANT: Merot, Bertrand  
APPLICANT: Dieryck, Wilfrid  
APPLICANT: Lenee, Philippe  
APPLICANT: Marden, Michael  
APPLICANT: Gruber, Veronique  
APPLICANT: Pagnier, Renee-Jossee  
APPLICANT: Baudino, Sylvie  
APPLICANT: Poyart, Claude  
TITLE OF INVENTION: METHOD FOR PRODUCING HAEMIN PROTEINS USING PLANT  
TITLE OF INVENTION: CELLS.  
FILE REFERENCE: 8076.147USMO  
CURRENT APPLICATION NUMBER: US/10/085,853  
CURRENT FILING DATE: 2001-10-18  
PRIOR APPLICATION NUMBER: 08/983,564  
PRIOR FILING DATE: 1998-06-09  
PRIOR APPLICATION NUMBER: PCT/FR96/01123  
PRIOR FILING DATE: 1996-07-17  
PRIOR APPLICATION NUMBER: 95/08615  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 24  
LENGTH: 12  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Homo sapiens  
US-10-085-853-24

Alignment Scores:  
Pred. No.: 190  
Score: 20.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 13  
Gaps: 0

US-09-696-872-37 (1-4) x US-10-085-853-24 (1-12)

QY 1 LysAspGluLeu 4  
|||  
1 AAAGATGAGCTA 12

Db

RESULT 3  
US-10-127-427-1  
Sequence 1, Application US/10127427  
Publication No. US20030051275A1  
GENERAL INFORMATION:  
APPLICANT: Paul CHRISTOU, Eva STROGER, Rainer FISCHER, Carmen MARTIN-VAQUERO,  
Stefan SC  
TITLE OF INVENTION: METHODS AND MEANS FOR EXPRESSION OF MAMMALIAN  
POLYPEPTIDES  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski L.L.P.  
STREET: 666 Fifth Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/127,427  
FILING DATE: 23-Apr-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/333,527  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 60/089,322  
FILING DATE: June 15, 1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary Anne Schofield  
REGISTRATION NUMBER: 36,669  
REFERENCE/DOCKET NUMBER: KL/JIC 202.1 - JEL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 318-3000  
TELEFAX: (212) 752-5958  
582015  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-10-127-427-1

Alignment Scores:  
Pred. No.: 190  
Score: 20.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 14  
Gaps: 0

US-09-696-872-37 (1-4) x US-10-127-427-1 (1-12)

QY 1 LysAspGluLeu 4  
|||  
1 AAAGATGAGCTC 12

Db

RESULT 4  
US-10-287-919-1514/C  
Sequence 1514, Application US/10287919  
Publication No. US20030085830A1

```
; GENERAL INFORMATION:
; APPLICANT: Feldmann, Richard J.; Global Determinants, Inc.
; TITLE OF INVENTION: Methanococcus jannaschii complete genome.
; FILE REFERENCE: Jim Zegeer Law Offices - 703-684-8333
; CURRENT FILING DATE: 2002-11-05
; NUMBER OF SEQ ID NOS: 2706
; SOFTWARE: Proprietary
; SEQ ID NO 1514
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii complete genome.
; FEATURE:
; LOCATION: (809601)...(809615)
; OTHER INFORMATION: Chromosome = 1 Strand = negative ConnectorObjectNumber = 1910
US-10-287-919-1514

Alignment Scores:
Pred. No.: 237 Length: 15
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-696-872-37 (1-4) x US-10-287-919-1514 (1-15)

QY 1 LysAspGluLeu 4
Db 12 AAGGATGAACCTA 1

RESULT 5
US-10-287-919-1515
; Sequence 1515, Application US/10287919
; Publication No. US20030085830A1
; GENERAL INFORMATION:
; APPLICANT: Feldmann, Richard J.; Global Determinants, Inc.
; TITLE OF INVENTION: Methanococcus jannaschii complete genome.
; FILE REFERENCE: Jim Zegeer Law Offices - 703-684-8333
; CURRENT APPLICATION NUMBER: US/10/287,919
; CURRENT FILING DATE: 2002-11-05
; NUMBER OF SEQ ID NOS: 2706
; SOFTWARE: Proprietary
; SEQ ID NO 1515
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii complete genome.
; FEATURE:
; LOCATION: (809601)...(809615)
; OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectorObjectNumber = 1905
US-10-287-919-1515

Alignment Scores:
Pred. No.: 237 Length: 15
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-696-872-37 (1-4) x US-10-287-919-1515 (1-15)

QY 1 LysAspGluLeu 4
Db 4 AAGGATGAACCTA 15

RESULT 6
US-10-287-919-2042
; Sequence 2042, Application US/10287919
; Publication No. US20030085830A1
; GENERAL INFORMATION:
; APPLICANT: Feldmann, Richard J.; Global Determinants, Inc.
; TITLE OF INVENTION: Methanococcus jannaschii complete genome.
```

```
; FILE REFERENCE: Jim Zegeer Law Offices - 703-684-8333
; CURRENT APPLICATION NUMBER: US/10/287,919
; CURRENT FILING DATE: 2002-11-05
; NUMBER OF SEQ ID NOS: 2706
; SOFTWARE: Proprietary
; SEQ ID NO 2042
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii complete genome.
; FEATURE:
; LOCATION: (1242502)...(1242516)
; OTHER INFORMATION: Chromosome = 1 Strand = positive ConnectorObjectNumber = 2612
US-10-287-919-2042

Alignment Scores:
Pred. No.: 237 Length: 15
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-696-872-37 (1-4) x US-10-287-919-2042 (1-15)

QY 1 LysAspGluLeu 4
Db 4 AAGGATGAACCTA 15

RESULT 7
US-10-287-919-2043/c
; Sequence 2043, Application US/10287919
; Publication No. US20030085830A1
; GENERAL INFORMATION:
; APPLICANT: Feldmann, Richard J.; Global Determinants, Inc.
; TITLE OF INVENTION: Methanococcus jannaschii complete genome.
; FILE REFERENCE: Jim Zegeer Law Offices - 703-684-8333
; CURRENT APPLICATION NUMBER: US/10/287,919
; CURRENT FILING DATE: 2002-11-05
; NUMBER OF SEQ ID NOS: 2706
; SOFTWARE: Proprietary
; SEQ ID NO 2043
; LENGTH: 15
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii complete genome.
; FEATURE:
; LOCATION: (1242502)...(1242516)
; OTHER INFORMATION: Chromosome = 1 Strand = negative ConnectorObjectNumber = 2611
US-10-287-919-2043

Alignment Scores:
Pred. No.: 237 Length: 15
Score: 20.00 Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-696-872-37 (1-4) x US-10-287-919-2043 (1-15)

QY 1 LysAspGluLeu 4
Db 12 AAGGATGAACCTA 1

RESULT 8
US-09-780-533A-652/c
; Sequence 652, Application US/09780533A
; Publication No. US20030060611A1
; GENERAL INFORMATION:
; APPLICANT: Ribozyne Pharmaceuticals, Inc.
; APPLICANT: Blatt, Larry
; APPLICANT: McSwiggen, Jim
; APPLICANT: Chowwira, Bharat
; APPLICANT: Haeblerli, Pete
```

```
/ TITLE OF INVENTION: Method and Reagent for the Inhibition of NOGO Gene
/ FILE REFERENCE: MBH00,878-A (400/011)
/ CURRENT APPLICATION NUMBER: US/09/780,533A
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: US 60/181,797
/ NUMBER OF SEQ ID NOS: 6679
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 652
/ LENGTH: 17
/ TYPE: RNA
/ ORGANISM: Homo sapiens
US-09-780-533A-652

Alignment Scores:
Pred. No.: 269          Length: 17
Score: 20.00           Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11                Gaps: 0

US-09-696-872-37 (1-4) x US-09-780-533A-652 (1-17)

QY 1 LysApGjLeu 4
DB 16 AAAGATGAATC 5

RESULT 9
US-09-780-533A-653/c
/ Sequence 653, Application US/09780533A
/ Publication No. US20030060611A1
/ GENERAL INFORMATION:
/ APPLICANT: Ribozyme Pharmaceuticals, Inc.
/ APPLICANT: Blatt, Larry
/ APPLICANT: McSwiggen, Jim
/ APPLICANT: Chowrita, Bharat
/ APPLICANT: Haeblerli, Pete
/ TITLE OF INVENTION: Method and Reagent for the Inhibition of NOGO Gene
/ FILE REFERENCE: MBH00,878-A (400/011)
/ CURRENT APPLICATION NUMBER: US/09/780,533A
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: US 60/181,797
/ NUMBER OF SEQ ID NOS: 6679
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 653
/ LENGTH: 17
/ TYPE: RNA
/ ORGANISM: Homo sapiens
US-09-780-533A-653

Alignment Scores:
Pred. No.: 269          Length: 17
Score: 20.00           Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11                Gaps: 0

US-09-696-872-37 (1-4) x US-09-780-533A-653 (1-17)

QY 1 LysApGjLeu 4
DB 15 AAAGATGAATC 4

RESULT 10
US-09-780-533A-654/c
/ Sequence 654, Application US/09780533A
/ Publication No. US20030060611A1
/ GENERAL INFORMATION:
/ APPLICANT: Ribozyme Pharmaceuticals, Inc.
/ APPLICANT: Blatt, Larry
```

```
/ APPLICANT: McSwiggen, Jim
/ APPLICANT: Chowrita, Bharat
/ APPLICANT: Haeblerli, Pete
/ TITLE OF INVENTION: Method and Reagent for the Inhibition of NOGO Gene
/ FILE REFERENCE: MBH00,878-A (400/011)
/ CURRENT APPLICATION NUMBER: US/09/780,533A
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: US 60/181,797
/ NUMBER OF SEQ ID NOS: 6679
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 654
/ LENGTH: 17
/ TYPE: RNA
/ ORGANISM: Homo sapiens
US-09-780-533A-654

Alignment Scores:
Pred. No.: 269          Length: 17
Score: 20.00           Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11                Gaps: 0

US-09-696-872-37 (1-4) x US-09-780-533A-654 (1-17)

QY 1 LysApGjLeu 4
DB 12 AAAGATGAATC 1

RESULT 11
US-09-780-533A-1501/c
/ Sequence 1501, Application US/09780533A
/ Publication No. US20030060611A1
/ GENERAL INFORMATION:
/ APPLICANT: Ribozyme Pharmaceuticals, Inc.
/ APPLICANT: Blatt, Larry
/ APPLICANT: McSwiggen, Jim
/ APPLICANT: Chowrita, Bharat
/ APPLICANT: Haeblerli, Pete
/ TITLE OF INVENTION: Method and Reagent for the Inhibition of NOGO Gene
/ FILE REFERENCE: MBH00,878-A (400/011)
/ CURRENT APPLICATION NUMBER: US/09/780,533A
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: US 60/181,797
/ NUMBER OF SEQ ID NOS: 6679
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 1501
/ LENGTH: 17
/ TYPE: RNA
/ ORGANISM: Homo sapiens
US-09-780-533A-1501

Alignment Scores:
Pred. No.: 269          Length: 17
Score: 20.00           Matches: 4
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11                Gaps: 0

US-09-696-872-37 (1-4) x US-09-780-533A-1501 (1-17)

QY 1 LysApGjLeu 4
DB 14 AAAGATGAATC 3

RESULT 12
US-09-969-373-3751
/ Sequence 3751, Application US/09969373
/ Patent No. US20020133852A1
```

GENERAL INFORMATION:  
APPLICANT: Effertz, Roger J.  
APPLICANT: Haugse, Brian M.  
TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping  
FILE REFERENCE: 38-10(52679)A  
CURRENT APPLICATION NUMBER: US/09/969,373  
PRIOR FILING DATE: 2001-10-02  
PRIOR APPLICATION NUMBER: US 09/754,853  
PRIOR FILING DATE: 2001-01-05  
PRIOR APPLICATION NUMBER: US 09/760,427  
PRIOR FILING DATE: 2001-01-13  
PRIOR APPLICATION NUMBER: US 09/855,768  
PRIOR FILING DATE: 2001-05-15  
NUMBER OF SEQ ID NOS: 4593  
SEQ ID NO 3751  
LENGTH: 19  
TYPE: DNA  
ORGANISM: Glycine max  
US-09-969-373-3751

Alignment Scores:  
Pred. No.: 301 Length: 19  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 10 Gaps: 0

US-09-696-872-37 (1-4) x US-09-969-373-3751 (1-19)

OY 1 LysApGluLeu 4  
Db 5 AAGGCGAATTG 16

RESULT 13  
US-09-752-983-229  
Sequence 229, Application US/09752983  
Patent No. US20010016575A1  
GENERAL INFORMATION:  
APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.  
APPLICANT: Graham, Brett P. Monia  
TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN MDN2  
TITLE OF INVENTION: EXPRESSION  
NUMBER OF SEQUENCES: 271  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Jane Massey Licata  
STREET: 66 East Main Street  
CITY: Marilton  
STATE: NJ  
COUNTRY: U.S.A.  
ZIP: 08053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM PC  
OPERATING SYSTEM: WINDOWS 95  
SOFTWARE: WORDPERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/752,983  
FILING DATE: 02-Jan-2001  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/280,805  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Licata, Jane Massey  
REGISTRATION NUMBER: 32,257  
REFERENCE/DOCKET NUMBER: ISPH-0346  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 609-810-1454  
TELEFAX: 609-810-1454  
INFORMATION FOR SEQ ID NO: 229:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 base pairs

TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
ANTI-SENSE: Yes  
US-09-752-983-229

Alignment Scores:  
Pred. No.: 317 Length: 20  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x US-09-752-983-229 (1-20)

OY 1 LysApGluLeu 4  
Db 2 AAGGATGAGCTA 13

RESULT 14  
US-09-752-983-230  
Sequence 230, Application US/09752983  
Patent No. US20010016575A1  
GENERAL INFORMATION:  
APPLICANT: Loren J. Miraglia, Pamela Nero, Mark J.  
APPLICANT: Graham, Brett P. Monia  
TITLE OF INVENTION: ANTISENSE MODULATION OF HUMAN MDN2  
TITLE OF INVENTION: EXPRESSION  
NUMBER OF SEQUENCES: 271  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Law Offices of Jane Massey Licata  
STREET: 66 East Main Street  
CITY: Marilton  
STATE: NJ  
COUNTRY: U.S.A.  
ZIP: 08053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM PC  
OPERATING SYSTEM: WINDOWS 95  
SOFTWARE: WORDPERFECT 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/752,983  
FILING DATE: 02-Jan-2001  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/280,805  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Licata, Jane Massey  
REGISTRATION NUMBER: 32,257  
REFERENCE/DOCKET NUMBER: ISPH-0346  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 609-810-1454  
TELEFAX: 609-810-1454  
INFORMATION FOR SEQ ID NO: 230:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
ANTI-SENSE: Yes  
US-09-752-983-230

Alignment Scores:  
Pred. No.: 317 Length: 20  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x US-09-752-983-230 (1-20)

Qy 1 Lysaapgiuleu 4  
|||||  
Db 6 AAGATGAGCTA 17

## RESULT 15

US-09-791-406-66/c  
; Sequence 66, Application US/09791406  
; Patent No. US20020147165A1  
; GENERAL INFORMATION:  
; APPLICANT: C. Frank Bennett  
; APPLICANT: Robert Rothlein  
; APPLICANT: Takashi Kei Kishimoto  
; APPLICANT: Lex M. Cowsett  
; TITLE OF INVENTION: ANTISENSE MODULATION OF CALRETICULIN EXPRESSION  
; FILE REFERENCE: RPS-0097  
; CURRENT APPLICATION NUMBER: US/09/791,406  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 89  
; SEQ ID NO 66  
; LENGTH: 20  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Antisense Oligonucleotide  
US-09-791-406-66

## Alignment Scores:

Pred. No.:	317	Length:	20
Score:	20.00	Matches:	4
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-696-872-37 (1-4) x US-09-791-406-66 (1-20)

Qy 1 Lysaapgiuleu 4  
|||||  
Db 13 AAGGACGAGCTG 2

Search completed: August 26, 2003, 21:02:50  
Job time : 12.504 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 26, 2003, 16:41:45 ; Search time 107.894 Seconds  
(without alignments)  
901.052 Million cell updates/sec

Title: US-09-696-872-37  
Perfect score: 20  
Sequence: 1 KDEL 4

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2281392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgm2\_1/USPTO\_epool/US09696872/runat\_26082003\_151137\_3236/app\_query.fasta\_1.462  
-DB=EST -QFMT=fastp -SUFFIX=rat -MINMATCH=0.1 -DOORCL=0 -DOOREXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09696872\_@CGN\_1\_1\_3586\_@runat\_26082003\_151137\_3236 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGOQUERY -NEG SCORES=0 -WAIT -DISPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_hcc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hcc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fut:\*  
22: em\_gss\_mus:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	* Query Match	Length	DB	ID	Description
1	20	100.0	25	9	AU254198	AU254198 AU254198
2	20	100.0	32	14	D19573	D19573 MUSGS00978
3	20	100.0	33	9	AU254805	AU254805 AU254805
4	20	100.0	38	9	AL801834	AL801834 AL801834
5	20	100.0	49	9	A1130296	A1130296 SMOVL3CAN
6	20	100.0	51	9	A1208498	A1208498 qgs5h03.x
7	20	100.0	51	29	B2766501	B2766501 SALK_1374
8	20	100.0	52	9	AV555282	AV555282 AV555282
9	20	100.0	52	28	BH620522	BH620522 1007066F1
10	20	100.0	53	9	AA837889	AA837889 oea4c01.b
11	20	100.0	56	9	AU259239	AU259239 AU259239
12	20	100.0	59	29	B2769592	B2769592 SALK_1424
13	20	100.0	60	12	BJ080266	BJ080266 BJ080266
14	20	100.0	60	28	BH011469	BH011469 BG02177-5
15	20	100.0	61	9	AV533632	AV533632 AV533632
16	20	100.0	61	29	AL942430	AL942430 Arabidops
17	20	100.0	62	14	T57059	T57059 yc20a01.81
18	20	100.0	63	28	BH857575	BH857575 SALK_0170
19	20	100.0	64	9	AM420167	AM420167 fj88b03.y
20	20	100.0	64	12	B1142411	B1142411 SMOV3WCAM
21	20	100.0	64	12	B1142469	B1142469 SMOV3WCAM
22	20	100.0	65	14	CA840831	CA840831 pf60a03.y
23	20	100.0	66	13	BO519722	BO519722 r439h08.y
24	20	100.0	67	9	AA519033	AA519033 v126d01.r
25	20	100.0	69	28	BH856618	BH856618 SALK_0794
26	20	100.0	70	28	BH856626	BH856626 SALK_0794
27	20	100.0	70	28	BH856627	BH856627 SALK_0794
28	20	100.0	70	28	BH856633	BH856633 SALK_0794
29	20	100.0	70	28	BH856634	BH856634 SALK_0794
30	20	100.0	70	28	BH856640	BH856640 SALK_0793
31	20	100.0	71	9	AV533656	AV533656 AV533656
32	20	100.0	71	13	BU672589	BU672589 haeg002xj
33	20	100.0	72	9	AW542189	AW542189 C0154F09-3
34	20	100.0	73	10	BE420479	BE420479 SMOVL2CAS
35	20	100.0	73	14	CB832348	CB832348 SWbmtCAV
36	20	100.0	73	14	CB832443	CB832443 SWbmtCAV
37	20	100.0	73	14	CB832487	CB832487 SWbmtCAV
38	20	100.0	73	14	CB832565	CB832565 SWbmtCAV
39	20	100.0	74	13	BO613177	BO613177 r402d06.y
40	20	100.0	77	9	A1952392	A1952392 wx73b08.x
41	20	100.0	78	12	B1065092	B1065092 pglf1n.pk0
42	20	100.0	78	12	B1468738	B1468738 sa1002506.f
43	20	100.0	78	28	AZ310739	AZ310739 IM0025P15
44	20	100.0	79	10	BE316173	BE316173 NF033C07L
45	20	100.0	79	28	AZ614196	AZ614196 IM0442124

#### ALIGNMENTS

RESULT 1  
LOCUS AU254198 25 bp mRNA linear EST 25-APR-2002  
DEFINITION AU254198 3'-directed mouse cDNA library Mus musculus cDNA clone  
ACCESSION AU254198  
VERSION AU254198  
KEYWORDS AU254198.1 GI:20315754  
SOURCE EST.  
ORGANISM Mus musculus (house mouse)  
MUS musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 25)

AUTHORS Kato, K. and Matoba, R.  
TITLE Generation of expressed sequence tags from mouse brain  
JOURNAL Unpublished  
COMMENT Contact: Kikuya Kato  
Graduate School of Biological Sciences  
Nara Institute of Science and Technology  
8916-5 Takayama, Ikoma, Nara 630-0101, Japan  
Tel: 81-743-72-5581  
Fax: 81-743-72-5589  
Email: kkatoc@bs.aist-nara.ac.jp,  
URL: <http://love2.aist-nara.ac.jp/BEI/index.html>.

## FEATURES

## source

1.25  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="BED0000932"  
/tissue\_type="brain"  
/clone\_lib="3'-directed mouse cDNA library"

BASE COUNT 13 a 3 c 4 g 5 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.53e+03 Length: 25  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AU254198 (1-25)

Qy 1 LysaepgJuleu 4  
|||||  
Db 10 AAAGATGAATTG 21

## RESULT 2

D19573 32 bp mRNA linear EST 12-DEC-1995  
LOCUS MUSGS00978 Mouse 3'-directed Mus musculus domesticus cDNA clone  
DEFINITION mb0732 3', mRNA sequence.  
D19573  
ACCESSION D19573.1 GI:1089429  
VERSION  
KEYWORDS EST.  
SOURCE Mus musculus domesticus (western European house mouse)  
ORGANISM Mus musculus domesticus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 32)  
AUTHORS Kawamoto, S., Okubo, K., Yoshii, J., Katsuki, M. and Matsubara, K.  
TITLE Analysis of gene expression in mouse embryogenesis by 3'-directed  
CDNA sequencing  
JOURNAL Unpublished  
COMMENT Contact: Kawamoto, S., Okubo, K., Yoshii, J., Katsuki, M. and Matsubara, K.  
Institute for Cellular and Molecular Biology  
Osaka University  
3-1 Yamada-oka, Suita, Osaka 565, Japan.  
location/Qualifiers

1.32  
/organism="Mus musculus domesticus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10092"  
/clone="mb0732"  
/tissue\_type="decidual tissue (day 6.5-8.5 of gestation)"  
/clone\_lib="mouse 3'-directed"  
BASE COUNT 14 a 4 c 5 g 9 t  
ORIGIN

## Alignment Scores:

Pred. No.: 2.14e+03 Length: 32  
Score: 20.00 Matches: 4

Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 14 Gaps: 0

US-09-696-872-37 (1-4) x D19573 (1-32)

Qy 1 LysaepgJuleu 4  
|||||  
Db 10 AAAGATGAATTG 21

## RESULT 3

AU254805 33 bp mRNA linear EST 25-APR-2002  
LOCUS AU254805 3'-directed mouse cDNA library Mus musculus cDNA clone  
DEFINITION BED000418 3', mRNA sequence.  
AU254805  
ACCESSION AU254805.1 GI:20316936  
VERSION  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 33)  
AUTHORS Kato, K. and Matoba, R.  
TITLE Generation of expressed sequence tags from mouse brain  
JOURNAL Unpublished  
COMMENT Contact: Kikuya Kato  
Graduate School of Biological Sciences  
Nara Institute of Science and Technology  
8916-5 Takayama, Ikoma, Nara 630-0101, Japan  
Tel: 81-743-72-5581  
Fax: 81-743-72-5589  
Email: kkatoc@bs.aist-nara.ac.jp,  
URL: <http://love2.aist-nara.ac.jp/BEI/index.html>.

1.33  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="BED000418"  
/tissue\_type="brain"  
/clone\_lib="3'-directed mouse cDNA library"

BASE COUNT 17 a 3 c 4 g 9 t  
ORIGIN

## Alignment Scores:

Pred. No.: 2.23e+03 Length: 33  
Score: 20.00 Matches: 4  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AU254805 (1-33)

Qy 1 LysaepgJuleu 4  
|||||  
Db 10 AAAGATGAATTG 21

## RESULT 4

AL801834 38 bp mRNA linear EST 27-JUN-2002  
LOCUS AL801834 XGC-neurula Silurana tropicalis cDNA clone TNeu097103 5',  
DEFINITION mRNA sequence.  
AL801834  
ACCESSION AL801834.1 GI:21588202  
VERSION  
KEYWORDS EST.  
SOURCE Silurana tropicalis (western clawed frog)  
ORGANISM Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Silurana.

1.32  
/organism="Silurana tropicalis"  
/mol\_type="mRNA"  
/strain="XGC-neurula"  
/db\_xref="taxon:7955"  
/clone="TNeu097103"  
/tissue\_type="neural tissue (stage 55)"  
/clone\_lib="cDNA library"

BASE COUNT 14 a 4 c 5 g 9 t  
ORIGIN

REFERENCE 1 (bases 1 to 38)  
 AUTHORS Taylor, R., Ashurst, J.L., Croning, M.D.R., Zorn, A.M. and Rogers, J.  
 TITLE Sanger Xenopus tropicalis EST project 2002  
 JOURNAL Unpublished  
 COMMENT Contact: Taylor R  
 Sanger Centre  
 Hinxton, Cambridgeshire, CB10 1SA, UK  
 Email: trop@sanger.ac.uk  
 Sanger Xenopus tropicalis EST project 2001  
 TROPICALIS\_SEQUENCE\_ID: TNeu097103.p1csp6  
 Sequencing primer: F1CSP6  
 This sequence is from a Xenopus Gene Collection (XGC) library  
 constructed by Aaron M. Zorn.  
 location/Qualifiers  
 1..38  
 /organism="Silurana tropicalis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8364"  
 /clone="TNeu097103"  
 /dev\_stage="neurula"  
 /lab\_host="Escherichia coli DH10B"  
 /note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA  
 was oligo dt primed from 5' end of poly A+ RNA from neurula.  
 EcoRI-NotI cut cDNA was then ligated into pCS107 with  
 EcoRI at the 5' end and NotI at the 3' end."  
 BASE COUNT 11 a 9 c 9 g 9 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.7e+03 Length: 38  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0

US-09-696-872-37 (1-4) x AL601834 (1-38)

OY 1 LysApGtGluLeu 4  
 Db 16 AAGAGCGAAGCTA 27

RESULT 5  
 A1130296 49 bp mRNA linear EST 26-MAR-1999  
 LOCUS SWOVL3CAN08G12 Onchocerca volvulus infective larva cDNA  
 DEFINITION (SWM94ML-OVL3) Onchocerca volvulus cDNA clone onch276 5' similar to  
 TR:001657 001657 FRUCTOSE-BISPHOSPHATE ALDOLASE ;, mRNA sequence.  
 A1130296  
 VERSION A1130296.1 GI:3600314  
 KEYWORDS EST.  
 SOURCE Onchocerca volvulus  
 ORGANISM Onchocerca volvulus  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Sphurida; Filarioidea;  
 Onchocercidae; Onchocerca.  
 1 (bases 1 to 49)  
 Williams, S.A., Lizotte-Maniewski, M., Laney, S., Wenhong, L., Hillier  
 L., Allen, M., Bowles, L., Geisel, S., Jost, S., Kucaba, T., Martin, J.,  
 Steptoe, M., Theising, B., White, Y., Wylie, T., Chappell, J., Pearson, B.,  
 Gibbons, M., Harvey, N., Page, D., Chamberlain, A., Morales, R.,  
 Schurr, R., Rittler, E., Kohn, S., Underwood, K. and Maira, M.  
 Molecular Parasitology OVL3  
 Unpublished  
 Contact: Steven A. Williams  
 Molecular Parasitology  
 Smith College Department of Biological Sciences  
 Department of Biological Sciences, Clark Science Center, Smith  
 College, Northampton, MA, 01063, USA  
 Tel: 4135853826  
 Fax: 4135853786  
 Email: genome@smith.edu  
 The library was constructed by Wenhong Lu. The library is available

from Dr. S.A. Williams, email genome@smith.edu when requesting this  
 clone from Dr. Williams, please reference the Williams lab clone id  
 - SWOVL3CAN08G12  
 Trace considered overall poor quality  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: T3 ET from Amerisham  
 High quality sequence stop: 1.  
 location/Qualifiers  
 1..49  
 /organism="Onchocerca volvulus"  
 /mol\_type="mRNA"  
 /strain="Sierra Leone"  
 /db\_xref="taxon:6282"  
 /clone="onch276"  
 /lab\_host="XLI-Blue MRF"  
 /clone\_lib="Onchocerca volvulus infective larva cDNA  
 (SWM94ML-OVL3)"  
 /note="Vector: lambda Unizap XR; Site\_1: EcoR I; Site\_2:  
 Xho I; Cutaneous filarial nematode parasite of humans.  
 mRNA was prepared from third stage infective larvae of  
 Onchocerca volvulus isolated from mosquitoes 10 days after  
 infection and converted to double stranded cDNA using  
 reverse transcriptase and oligo(dt) followed by RNase H  
 and DNase I. The library had 1.8 x 10E5 independent  
 recombinants and average insert size was 900 base pairs.  
 The library was constructed by Wenhong Lu. The library is  
 available from Dr. S.A. Williams, email genome@smith.edu."  
 BASE COUNT 15 a 9 c 11 g 14 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.8e+03 Length: 49  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0

US-09-696-872-37 (1-4) x A1130296 (1-49)

OY 1 LysApGtGluLeu 4  
 Db 26 AAAGATGATG 37

RESULT 6  
 A1208498 51 bp mRNA linear EST 30-NOV-1998  
 LOCUS qg5th03.x1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1839125  
 DEFINITION qg5th03.x1 Soares testis NHT Homo sapiens cDNA clone IMAGE:1839125  
 3' similar to gb:U73628 KAPPA CASEIN PRECURSOR (HUMAN);, mRNA  
 sequence.  
 A1208498  
 VERSION A1208498.1 GI:3770440  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 51)  
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo  
 , Ph.D.  
 cDNA library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bdrip/image/image.html  
 Trace considered overall poor quality

Insert Length: 761 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence scop: 1.  
 Location/Qualifiers  
 1..51

source  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1839125"  
 /sex="male"  
 /lab\_host="DH10B"  
 /clone\_lib="Soares testis NT"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I, Site 2: Eco RI, 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer (5' TGTTACATCTGAGTGGAGCGCCGCCAATTTTCTTTT 3'). Double-stranded cDNA was ligated to Eco RI adaptor (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cots, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 21 a 10 c 7 g 13 t

ORIGIN

Alignment Scores:  
 Pred. No.: 4.01e+03 Length: 51  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Gaps: 0

US-09-696-872-37 (1-4) x A1208498 (1-51)

QY 1 Lysaapglu4  
 |||||  
 8 AAGATGAACCTA 19

LOCUS B2766501 51 bp DNA linear GSS 13-MAR-2003  
 DEFINITION SALK\_137492.43.00 x Arabidopsis thaliana TDNA insertion line  
 Arabidopsis thaliana genomic clone SALK\_137492.43.00.x, genomic survey sequence.

ACCESSION B2766501  
 VERSION B2766501  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana

REFERENCE Alonso,J.M., Leese,T.J., Barajas,P., Chen,H., Cheuk,R., Gadinab,  
 C., Jeske,A., Karner,M., Kim,C.J., Parker,H., Prednis,L., Shin,P.,  
 Zimmerman,J., and Ecker,J.R.  
 A Sequence-Indexed Library of Insertion Mutations in the  
 Arabidopsis Genome

TITLE Unpublished  
 JOURNAL Contact: Joseph R. Ecker  
 COMMENT Salt Institute Genomic Analysis Laboratory (SIGNAL)  
 The Salt Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6379  
 Email: ecker@sal.k.edu

FEATURES  
 source  
 This is single pass sequence recovered from the left border of  
 TDNA. This sequence lies within 300 bases of the 5' end of  
 At1g20200 and 300 bases of the 5' end of At1g20210.  
 Class: TDNA tagged.  
 Location/Qualifiers  
 1..51

/organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone="SALK\_137492.43.00.x"  
 /clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at [http://signal.sal.k.edu/tdna\\_protocols.html](http://signal.sal.k.edu/tdna_protocols.html)"

BASE COUNT 10 a 18 c 8 g 15 t

ORIGIN

Alignment Scores:  
 Pred. No.: 4.01e+03 Length: 51  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Gaps: 0

US-09-696-872-37 (1-4) x B2766501 (1-51)

QY 1 Lysaapglu4  
 |||||  
 16 AAGAGCGAGCTG 5

RESULT 8 AV555282 52 bp mRNA linear EST 06-SEP-2000  
 LOCUS AV555282 Arabidopsis thaliana green silique Columbia Arabidopsis  
 DEFINITION thaliana cDNA clone SQ010f02F 3', mRNA sequence.

ACCESSION AV555282  
 VERSION AV555282.1 GI:8726697  
 KEYWORDS EST.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana

REFERENCE Alonso,J.M., Leese,T.J., Barajas,P., Chen,H., Cheuk,R., Gadinab,  
 C., Jeske,A., Karner,M., Kim,C.J., Parker,H., Prednis,L., Shin,P.,  
 Zimmerman,J., and Ecker,J.R.  
 A Sequence-Indexed Library of Insertion Mutations in the  
 Arabidopsis Genome

TITLE Unpublished  
 JOURNAL Contact: Erika Asamizu  
 COMMENT The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: [asamizu@kazusa.or.jp](mailto:asamizu@kazusa.or.jp), URL: <http://www.kazusa.or.jp/en/plant/>.

FEATURES  
 source  
 1..52  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone="SQ010f02F"  
 /tissue\_type="green silique"  
 /clone\_lib="Arabidopsis thaliana green silique Columbia"  
 /note="Vector: pBluescript SK-, Site 1: EcoRI, Site 2: XhoI"

BASE COUNT 21 a 7 c 4 g 20 t

ORIGIN

Alignment Scores:  
 Pred. No.: 4.11e+03 Length: 52  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AV555282 (1-52)

QY 1 LysApGluLeu 4  
 DB 27 AAAGATGAGTTA 16

RESULT 9  
 LOCUS BH620522/c 52 bp DNA linear GSS 30-JAN-2002  
 DEFINITION 1007066F12.2EL x1 1007 - Rescuemu Grid H Zea mays genomic, genomic survey sequence.

ACCESSION BH620522.1 GI:18432496  
 VERSION BH620522  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 52)  
 Walbot,V.

JOURNAL Maize genomic sequences found using engineered Rescuemu transposon  
 COMMENT Unpublished  
 CONTACT: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu

FEATURES  
 source  
 Location/Qualifiers  
 1..52  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultivar="mixed background W23/A188/B73"  
 /db\_xref="taxon:4577"  
 /tissue\_type="leaf"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /clone\_lib="1007 - Rescuemu Grid H"  
 /note="Organ: leaf; Vector: Rescuemu (engineered from Bluescript backbone); Site 1: BamHI; Site 2: BglII; Rescuemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescuemu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'Rescuemu', 'Grid H was grown at Berkeley in 2001. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin.'"

BASE COUNT 10 a 18 c 9 g 15 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 4.11e+03 Length: 52  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 28 Gaps: 0

US-09-696-872-37 (1-4) x BH620522 (1-52)

QY 1 LysApGluLeu 4  
 DB 32 AAAGATGAGCTG 21

RESULT 10  
 LOCUS AA837889 53 bp mRNA linear EST 18-MAR-1998  
 DEFINITION o040c01.61 NCI CGAP Pr25 Homo sapiens CDNA clone IMAGE:1411008 similar to gb:M73628 KAPPA CASEIN PRECURSOR (HUMAN)"; mRNA sequence.

ACCESSION AA837889  
 VERSION AA837889.1 GI:2913546  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 53)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb@remail.nih.gov  
 Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright, Ph.D.

CDNA Library Preparation: Stratagene, Inc.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, at: www-bio.lnl.gov/biopr/image/image.html  
 Insert length: 857 Std Error: 0.00  
 Seq primer: -40m13 fwd, ET from Amersham  
 High quality sequence stop: 1.  
 Location/Qualifiers  
 1..53  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1411008"  
 /tissue\_type="epithelium (cell line)"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="NCI CGAP Pr25"  
 /note="Organ: prostate; Vector: Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Normal prostate epithelial cell line (HPV immortalized). 5' adaptor sequence: 5' GAATTCGGCAGG 3' 3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' Average insert size: 1.1 kb."

BASE COUNT 21 a 12 c 8 g 11 t 1 others  
 ORIGIN

## Alignment Scores:

Pred. No.: 4.22e+03 Length: 53  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AA837889 (1-53)

QY 1 LysApGluLeu 4  
 DB 8 AAAGATGAGCTA 19

RESULT 11  
 LOCUS AU259239/c 56 bp mRNA linear EST 25-APR-2002  
 DEFINITION AU259239 3'-directed mouse cDNA library Mus musculus cDNA clone BED0014799 3', mRNA sequence.  
 ACCESSION AU259239

VERSION AU259239.1 GI:20325573  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 56)  
 Kato, K. and Matsuda, R.  
 Generation of expressed sequence tags from mouse brain  
 Unpublished  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: Kikuya Kato  
 Graduate School of Biological Sciences  
 Nara Institute of Science and Technology  
 8916-5 Takayama, Ikoma, Nara 630-0101, Japan  
 Tel: 81-743-72-5581  
 Fax: 81-743-72-5589  
 Email: kkatoda@ist-nara.ac.jp,  
 URL: <http://love2.ist-nara.ac.jp/BED/index.html>.  
 Location/Qualifiers  
 1..56  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="BED0014799"  
 /issue\_type="Brain"  
 /clone\_lib="3'-directed mouse cDNA library"  
 BASE COUNT 24 a 10 c 7 g 14 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4.54e+03 Length: 56  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-696-872-37 (1-4) x AU259239 (1-56)  
 QY 1 LysAAGGtLleu 4  
 |||||  
 39 AAGATGAATG 28  
 RESULT 12 59 bp DNA linear GSS 13-MAR-2003  
 BZ769592  
 LOCUS SALK\_142419.31.10.n Arabidopsis thaliana TDNA insertion line  
 DEFINITION Arabidopsis thaliana genomic clone SALK\_142419.31.10.n, genomic survey sequence.  
 ACCESSION BZ769592  
 VERSION BZ769592.1 GI:28943276  
 KEYWORDS GSS.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 1 (bases 1 to 59)  
 Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karne, M., Kim, C.U., Parker, H., Prednis, L., Shim, P., Zimmerman, U., and Becker, J.R.  
 A Sequence-Indexed library of Insertion Mutations in the Arabidopsis Genome  
 Unpublished  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: Joseph R. Becker  
 Salk Institute Genomic Analysis Laboratory (SIGAL)  
 The Salk Institute for Biological Studies  
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
 Tel: 858 453 4100 x1752  
 Fax: 858 558 6379  
 Email: eckersalk.edu  
 This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated exon of Atlg03370.

FEATURES  
 source  
 Class: TDNA tagged.  
 Location/Qualifiers  
 1..59  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /strain="Columbia 0"  
 /db\_xref="taxon:3702"  
 /clone="SALK\_142419.31.10.n"  
 /clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at [http://signal.salk.edu/cdna\\_protocols.html](http://signal.salk.edu/cdna_protocols.html)"  
 BASE COUNT 17 a 9 c 14 g 19 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4.87e+03 Length: 59  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 29 Gaps: 0  
 US-09-696-872-37 (1-4) x BZ769592 (1-59)  
 QY 1 LysAAGGtLleu 4  
 |||||  
 37 AAGATGAATG 48  
 RESULT 13 60 bp mRNA linear EST 12-DEC-2001  
 BJ080266  
 LOCUS BJ080266  
 DEFINITION BJ080266 NIBB Mochi normalized Xenopus tailbud library Xenopus laevis cDNA clone XL075a15 3', mRNA sequence.  
 ACCESSION BJ080266  
 VERSION BJ080266.1 GI:17573878  
 KEYWORDS EST.  
 SOURCE Xenopus laevis (African clawed frog)  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Xenopus.  
 1 (bases 1 to 60)  
 Kitayama, A., Terasaka, C., Mochi, M., Ueno, N., Shin-i, T. and Kohara, Y.  
 Expressed genes in X. laevis embryo  
 Unpublished  
 TITLE  
 JOURNAL  
 COMMENT  
 Contact: Tadasu Shin-i  
 Center for Genetic Resource Information  
 National Institute of Genetics  
 1111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855  
 Email: tshin@gens.nig.ac.jp.  
 Location/Qualifiers  
 1..60  
 /organism="Xenopus laevis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8355"  
 /clone="XL075a15"  
 /tissue\_type="whole embryo"  
 /dev\_stage="stage 25"  
 /clone\_lib="NIBB Mochi normalized Xenopus tailbud library"  
 BASE COUNT 21 a 9 c 14 g 13 t 3 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 4.99e+03 Length: 60  
 Score: 20.00 Matches: 4

Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 100.00%  
 DB: 12 Gaps: 0

US-09-696-872-37 (1-4) x BU080266 (1-60)

OY 1 LysApGJuleu 4  
 DB 28 AAAGTGAAGCTG 39

RESULT 14  
 LOCUS BH011469  
 DEFINITION BG02177-5prine Drosophila melanogaster P[Gtl] P element insertion  
 lines Drosophila melanogaster genomic sequence recovered from 5'  
 end of P element, genomic survey sequence.

ACCESSION BH011469 60 bp DNA linear GSS 11-MAY-2001  
 VERSION BH011469  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 60)  
 Leyva, R., Hoskins, R., Liao, G., Morzen, N., Tsang, G., He, Y., Karpen  
 'G., Bellin, H., Rubin, G. and Spradling, A.,  
 TITLE The Berkeley Drosophila Genome Project Gene Disruption Project  
 JOURNAL Unpublished  
 COMMENT Contact: Gerald Rubin  
 Berkeley Drosophila Genome Project  
 University of California, Berkeley  
 LSA Building, Berkeley, CA 94720-3200, USA  
 Fax: 5106433947  
 Email: gerry@fruitfly.berkeley.edu

FEATURES  
 source  
 1..60  
 Location/Qualifiers  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone\_lib="Drosophila melanogaster P[Gtl] P element  
 insertion lines"  
 /note="inverse PCR was performed on Drosophila  
 melanogaster strains each of which contains one or more  
 P[Gtl] P-element transposon insertion. The resultant  
 fragment for each strain was directly sequenced to  
 determine the genomic sequence at the site of insertion.  
 Details of the protocols used can be found at  
 http://www.fruitfly.org/about/methods/inverse.pcr.html."

BASE COUNT 20 a 19 c 16 g 5 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4.99e+03 Length: 60  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 28 Gaps: 0

US-09-696-872-37 (1-4) x BH011469 (1-60)

OY 1 LysApGJuleu 4  
 DB 35 AAAGCGAAGCTC 46

RESULT 15  
 LOCUS AV533632 61 bp mRNA linear EST 01-SEP-2000  
 DEFINITION AV533632 Arabidopsis thaliana flower buds Columbia Arabidopsis  
 thaliana cDNA clone FB065C05F 3', mRNA sequence.

ACCESSION AV533632  
 VERSION AV533632  
 KEYWORDS AV533632.1 GI:8693915  
 EST.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 ; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 61)  
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.  
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation  
 of 12,028 non-redundant expressed sequence tags from normalized and  
 size-selected cDNA libraries  
 DNA Res. 7, 175-180 (2000)

JOURNAL MEDLINE  
 PUBMED 20363093  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
 source  
 1..61  
 Location/Qualifiers  
 /organism="Arabidopsis thaliana"  
 /mol\_type="mRNA"  
 /strain="Columbia"  
 /db\_xref="taxon:3702"  
 /clone\_lib="FB065C05F"  
 /tissue\_type="flower buds"  
 /clone\_lib="Arabidopsis thaliana flower buds Columbia"  
 /note="Vector: pBluescriptII SK-, Site\_1: EcoRI, Site\_2:  
 XhoI"

BASE COUNT 21 a 4 c 12 g 24 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5.1e+03 Length: 61  
 Score: 20.00 Matches: 4  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-696-872-37 (1-4) x AV533632 (1-61)

OY 1 LysApGJuleu 4  
 DB 9 AAAGTGAATG 20

Search completed: August 26, 2003, 19:21:15  
 Job time : 110.894 secs

**This Page Blank (uspio)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2003, 15:10:24 ; Search time 1733 Seconds

(without alignments)  
8781.524 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 372

Sequence: 1 aagcttaccatgaggaagta.....attccatcatgcctcgcag 372

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: GenEmbl:\*

2: gb\_ba:\*

3: gb\_hcg:\*

4: gb\_in:\*

5: gb\_om:\*

6: gb\_ov:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sbs:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sbs:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hcg\_hum:\*

31: em\_hcg\_inv:\*

32: em\_hcg\_other:\*

33: em\_hcg\_mus:\*

34: em\_hcg\_pln:\*

35: em\_hcg\_rtd:\*

36: em\_hcg\_mam:\*

37: em\_hcg\_vrt:\*

38: em\_ay:\*

39: em\_hggo\_hum:\*

40: em\_hggo\_mus:\*

41: em\_hggo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	372	100.0	372	6	AR121628
2	271.2	72.9	372	6	AR121631
3	264.8	71.2	372	6	AR121630
4	202	54.3	369	6	AR121627
5	194	52.2	369	6	AR121626
6	193.4	52.0	387	6	AR121623
7	192	51.6	387	6	AR121624
8	190.8	51.3	444	6	AR121632
9	177.6	47.7	357	6	AR121625
10	156.4	42.0	315	6	AR121629
11	142.6	38.3	2274	9	AB086984
12	142.6	38.3	2439	6	AX329932
13	142.6	38.3	2439	6	AX330698
14	142.6	38.3	2439	6	AX330917
15	142.6	38.3	2439	9	HUMCOMP
16	138	37.1	2710	9	BC033676
17	137.8	37.0	764	6	BD124595
18	137.8	37.0	764	6	BD126165
19	137.8	37.0	1779	9	BD127132
20	137.8	37.0	1779	9	AK074508
21	128.8	34.6	2438	10	AF033530
22	101.8	27.4	2438	10	AF033530
23	99.6	26.8	305	12	EVE291687
24	98.4	26.5	2421	10	RNCOMP
25	83.6	22.5	46275	9	AC003107
26	80	21.5	302	12	EVE291686
27	78.2	21.0	323	12	EVE291685
28	74	19.9	78	6	A92276
29	73.8	19.8	78	6	A92275
30	64.2	17.3	5719	10	MMCATS1
31	63.4	17.0	8923	10	AF318577
32	63.4	17.0	204302	2	AC068650
33	63.4	17.0	234787	2	AC073780
34	61	16.4	180286	10	AL645600
35	60.2	16.2	8524	10	AF257516
36	59.2	15.9	188880	10	AC093203
37	59.2	15.9	192939	2	AC112412
38	59.2	15.9	233181	2	AC111558
39	58.8	15.8	1440	14	BHV1UL7GN
40	58.6	15.8	225024	2	AC116415
41	58.4	15.7	161005	2	AC116595
42	58.4	15.7	172114	10	AC131903
43	58.2	15.6	261604	2	AC119819
44	58	15.6	3454	14	HAD293912
45	58	15.6	3454	14	HAD293913

# ALIGNMENTS

RESULT 1

LOCUS AR121628

DEFINITION Sequence 24 from patent US 6160088.

ACCESSION AR121628

VERSION AR121628.1 GI:14105204

KEYWORDS

SOURCE

ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 372)

AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.

TITLE KDEL receptor inhibitors

JOURNAL Patent: US 6160088-A 24 12-DEC-2000;

FEATURES

Location/Qualifiers

linear PAT 16-MAY-2001

```
source 1..372
/organism="unknown"
BASE COUNT 91 a 102 c 122 g 57 t
ORIGIN
Query Match 100.0%; Score 372; DB 6; Length 372;
Best Local Similarity 100.0%; Pred. No. 2.3e-67;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTACCATGGGAAGATGATGATTTTAAAGCTTCTGCGCCCTTGCGGCAAGTTCAGC 60
DB 1 AAGCTTACCATGGGAAGATGATGATTTTAAAGCTTCTGCGCCCTTGCGGCAAGTTCAGC 60
QY 61 GCTGCCAAAAGATTCAGACCTGGGTGAGACTGTTGTTCAAGACCTGGGCCCGCAGATG 120
DB 61 GCTGCCAAAAGATTCAGACCTGGGTGAGACTGTTGTTGAGGAGACTTTTAAACGGCAGTTTC 120
QY 121 CTTGGGGAATCTGCAGAAACCAACCGGCGCTGCAGAGAGTGCAGGACTGGCTGCGGCGAG 180
DB 121 CTTGGGGAATCTGCAGAAACCAACCGGCGCTGCAGAGAGTGCAGGACTGGCTGCGGCGAG 180
QY 181 CAGGTCAAGGAGATCACTTCTGAAAAAACAAGTGTAGTGTGACGCGGCGGCGGCGG 240
DB 181 CAGGTCAAGGAGATCACTTCTGAAAAAACAAGTGTAGTGTGACGCGGCGGCGGCGGCGG 240
QY 241 CAGCGCGACCGGAAACCGGACCGGACGCGGACGCGGACCGGAAACCGGACCGGAAACCG 300
DB 241 CAGCGCGACCGGAAACCGGACCGGACGCGGACGCGGACCGGAAACCGGACCGGAAACCG 300
QY 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATGTTGTAGCGCGCGCGAGATTTCAT 360
DB 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATGTTGTAGCGCGCGCGAGATTTCAT 360
QY 361 ATGCATCTCGAG 372
DB 361 ATGCATCTCGAG 372

RESULT 2
LOCUS AR121631 372 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 30 from patent US 6160088.
ACCESSION AR121631
VERSION AR121631.1 GI:14105207
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 372)
AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.
TITLE KDEL receptor inhibitors
JOURNAL Patent: US 6160088-A 30 12-DEC-2000;
FEATURES
source 1..372
/organism="unknown"
BASE COUNT 102 a 96 c 106 g 68 t
ORIGIN
Query Match 72.9%; Score 271.2; DB 6; Length 372;
Best Local Similarity 83.1%; Pred. No. 2.1e-46;
Matches 309; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 AAGCTTACCATGGGAAGATGATGATTTTAAAGCTTCTGCGCCCTTGCGGCAAGTTCAGC 60
DB 1 AAGCTTACCATGGGAAGATGATGATTTTAAAGCTTCTGCGCCCTTGCGGCAAGTTCAGC 60
QY 61 GCTGCCAAAAGATTCAGACCTGGGTGAGACTGTTGTTCAAGACCTGGGCCCGCAGATG 120
DB 61 GCTGCCAAAAGATTCAGACCTGGGTGAGACTGTTGTTGAGGAGACTTTTAAACGGCAGTTTC 120
QY 121 CTTGGGGAATCTGCAGAAACCAACCGGCGCTGCAGAGAGTGCAGGACTGGCTGCGGCGAG 180
DB 121 CTTGGGGAATCTGCAGAAACCAACCGGCGCTGCAGAGAGTGCAGGACTGGCTGCGGCGAG 180
QY 121 TTGGGTCAATATGACAACTTAAACAATCTCTGGGAGAGGTGAAGGACCTTCTGAGACAG 180
DB 121 TTGGGTCAATATGACAACTTAAACAATCTCTGGGAGAGGTGAAGGACCTTCTGAGACAG 180
```

```
QY 181 CAGGTCAAGGAGATCACTTCTGAAAAAACAAGTGTAGTGTGACGCGGCGGCGGCGG 240
DB 181 CAGGTCAAGGAGATCACTTCTGAAAAAACAAGTGTAGTGTGACGCGGCGGCGGCGGCGG 240
QY 241 CAGCGCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGACCGGAAACCG 300
DB 241 CAGCGCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGACCGGAAACCG 300
QY 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATGTTGTAGCGCGCGCGAGATTTCAT 360
DB 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATGTTGTAGCGCGCGCGAGATTTCAT 360
QY 361 ATGCATCTCGAG 372
DB 361 ATGCATCTCGAG 372

RESULT 3
LOCUS AR121630 372 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 28 from patent US 6160088.
ACCESSION AR121630
VERSION AR121630.1 GI:14105206
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 372)
AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.
TITLE KDEL receptor inhibitors
JOURNAL Patent: US 6160088-A 28 12-DEC-2000;
FEATURES
source 1..372
/organism="unknown"
BASE COUNT 96 a 105 c 109 g 62 t
ORIGIN
Query Match 71.2%; Score 264.8; DB 6; Length 372;
Best Local Similarity 82.0%; Pred. No. 4.5e-45;
Matches 305; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 AAGCTTACCATGGGAAGATGATGATTTTAAAGCTTCTGCGCCCTTGCGGCAAGTTCAGC 60
DB 1 AAGCTTACCATGGGAAGATGATGATTTTAAAGCTTCTGCGCCCTTGCGGCAAGTTCAGC 60
QY 61 GCTGCCAAAAGATTCAGACCTGGGTGAGACTGTTGTTCAAGACCTGGGCCCGCAGATG 120
DB 61 GCTGCCAAAAGATTCAGACCTGGGTGAGACTGTTGTTGAGGAGACTTTTAAACGGCAGTTTC 120
QY 121 CTTGGGGAATCTGCAGAAACCAACCGGCGCTGCAGAGAGTGCAGGACTGGCTGCGGCGAG 180
DB 121 CTTGGGGAATCTGCAGAAACCAACCGGCGCTGCAGAGAGTGCAGGACTGGCTGCGGCGAG 180
QY 181 CAGGTCAAGGAGATCACTTCTGAAAAAACAAGTGTAGTGTGACGCGGCGGCGGCGG 240
DB 181 CAGGTCAAGGAGATCACTTCTGAAAAAACAAGTGTAGTGTGACGCGGCGGCGGCGGCGG 240
QY 241 CAGCGCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGACCGGAAACCG 300
DB 241 CAGCGCGACCGGAAACCGGACCGGACCGGACCGGACCGGAAACCGGACCGGAAACCG 300
QY 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATGTTGTAGCGCGCGCGAGATTTCAT 360
DB 301 GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATGTTGTAGCGCGCGCGAGATTTCAT 360
QY 361 ATGCATCTCGAG 372
DB 361 ATGCATCTCGAG 372

RESULT 4
LOCUS AR121627
```

LOCUS AR121627 369 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 22 from patent US 6160088.  
ACCESSION AR121627  
VERSION AR121627.1 GI:14105203  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 369)  
AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.  
TITLE KDEL receptor inhibitors  
JOURNAL Patent: US 6160088-A 22 12-DEC-2000;  
FEATURES  
source Location/Qualifiers  
1..369  
/organism="unknown"  
BASE COUNT 94 a 94 c 121 g 60 t  
ORIGIN  
Query Match 54.3%; Score 202; DB 6; Length 369;  
Best Local Similarity 79.1%; Pred. No. 5,2e-32;  
Matches 239; Conservative 0; Mismatches 63; Indels 0; Gaps 0;  
QY 71 AAGATCCAGCTGGGTGAGACTGTTGTCAGACCTGGGCGCCGAGATGCTTCGGGAAC 130  
DB 68 AGGGATCCAGCTGGGTGAGACTGTTGTCAGACCTGGGCGCCGAGATGCTTCGGGAAC 127  
QY 131 TGCAGAAACCAACGCGCGCTGCGAGACGTCGGGGAAGTGGGAGAGAGTCAAGG 190  
DB 128 TACCCCAATGATGATGATGCTGGGAGAGCTCCGAGATGTCATGAGACGAGGTGAAG 187  
QY 191 AGATCACCTTCCTGTAACCAACGCTGATGAGTGTGACGCGTGGGCGCGAGCCGAGC 250  
DB 188 AGACCATGTTCTTGAGAAACACCATTTGCAAGATGCAAGCCCTGTGGCCGCGAGCCGAGC 247  
QY 251 CGAAACCCGAGCCGAGCGCGAGCGCGAGCCGAAACCGGAAACCGGAACCGGAAG 310  
DB 248 CGAAACCCGAGCCGAGCGCGAGCGCGAGCCGAAACCGGAAACCGGAACCGGAAG 307  
QY 311 GTACCGGATCATCAGAAAGATGATGTTAGGCGGCGCGAGAAATTCATATGATCTCG 370  
DB 308 GTACCGGATCATCAGAAAGATGATGTTAGGCGGCGCGAGAAATTCATATGATCTCG 367  
QY 371 AG 372  
DB 368 AG 369  
RESULT 5  
LOCUS AR121626 369 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 20 from patent US 6160088.  
ACCESSION AR121626  
VERSION AR121626.1 GI:14105202  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 369)  
AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.  
TITLE KDEL receptor inhibitors  
JOURNAL Patent: US 6160088-A 20 12-DEC-2000;  
FEATURES  
source Location/Qualifiers  
1..369  
/organism="unknown"  
BASE COUNT 87 a 103 c 120 g 59 t  
ORIGIN  
Query Match 52.2%; Score 194; DB 6; Length 369;  
Best Local Similarity 77.5%; Pred. No. 2,4e-30;  
Matches 234; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
QY 71 AAGATCCAGCTGGGTGAGACTGTTGTCAGACCTGGGCGCCGAGATGCTTCGGGAAC 130  
DB 68 AGGGATCCAGCTGGGTGAGACTGTTGTCAGACCTGGGCGCCGAGATGCTTCGGGAAC 127  
QY 131 TGCAGAAACCAACGCGCGCTGCGAGACGTCGGGGAAGTGGGAGAGAGTCAAGG 190  
DB 128 TACCCCAATGATGATGATGCTGGGAGAGCTCCGAGATGTCATGAGACGAGGTGAAG 187  
QY 191 AGATCACCTTCCTGTAACCAACGCTGATGAGTGTGACGCGTGGGCGCGAGCCGAGC 250  
DB 188 AGACCATGTTCTTGAGAAACACCATTTGCAAGATGCAAGCCCTGTGGCCGCGAGCCGAGC 247  
QY 251 CGAAACCCGAGCCGAGCGCGAGCGCGAGCCGAAACCGGAAACCGGAACCGGAAG 310  
DB 248 CGAAACCCGAGCCGAGCGCGAGCGCGAGCCGAAACCGGAAACCGGAACCGGAAG 307  
QY 311 GTACCGGATCATCAGAAAGATGATGTTAGGCGGCGCGAGAAATTCATATGATCTCG 370  
DB 308 GTACCGGATCATCAGAAAGATGATGTTAGGCGGCGCGAGAAATTCATATGATCTCG 367  
QY 371 AG 372  
DB 368 AG 369

DB 68 AGGGATCCAGCTGGGTGAGACTGTTGTCAGACCTGGGCGCCGAGATGCTTCGCCAGC 127  
QY 131 TGCAGAAACCAACGCGCGCTGCGAGACGTCGGGGAAGTGGGAGAGAGTCAAGG 190  
DB 128 TACCCCTTTCAACCAAGATCTCTAGTGAGCTTCGGGAGAGATCCGAGACAGGTGAAG 187  
QY 191 AGATCACCTTCCTGTAACCAACGCTGATGAGTGTGACGCGTGGGCGCGAGCCGAGC 250  
DB 188 AATGTCACTATCCGGAACACCATCATGAGATGTCAGGTGTGGGTCCGAGCCGAGC 247  
QY 251 CGAAACCCGAGCCGAGCGCGAGCGCGAGCCGAAACCGGAAACCGGAACCGGAAG 310  
DB 248 CGAAACCCGAGCCGAGCGCGAGCGCGAGCCGAAACCGGAAACCGGAACCGGAAG 307  
QY 311 GTACCGGATCATCAGAAAGATGATGTTAGGCGGCGCGAGAAATTCATATGATCTCG 370  
DB 308 GTACCGGATCATCAGAAAGATGATGTTAGGCGGCGCGAGAAATTCATATGATCTCG 367  
QY 371 AG 372  
DB 368 AG 369  
RESULT 6  
LOCUS AR121623 387 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 14 from patent US 6160088.  
ACCESSION AR121623  
VERSION AR121623.1 GI:14105199  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 387)  
AUTHORS Rothman,J.E., Mayhew,M. and Hoe,M.H.  
TITLE KDEL receptor inhibitors  
JOURNAL Patent: US 6160088-A 14 12-DEC-2000;  
FEATURES  
source Location/Qualifiers  
1..387  
/organism="unknown"  
BASE COUNT 94 a 110 c 125 g 58 t  
ORIGIN  
Query Match 52.0%; Score 193.4; DB 6; Length 387;  
Best Local Similarity 81.6%; Pred. No. 3,2e-30;  
Matches 249; Conservative 0; Mismatches 26; Indels 30; Gaps 1;  
QY 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAAGAAACCAACGCGCGCTGCAG 157  
DB 83 GTGGAGACTTGACCCCAACATGCTTGAGAAATCCAGAGATATATGCGGCGCTGCAG 142  
QY 158 ACGTGGGAGCTGGCTGGCGAGCAGATGTCAGGAGATACCTTCTGAAAAACGCTGA 217  
DB 143 ACGTGAGAGACTCTTGGCGAGCAGAGTCAAGAGATCACCTTCTGAAAGATACGCTGA 202  
QY 218 TGGAGTGTGAGCGCTGCG-----GGCCGAGCCGC 247  
DB 203 TGGATGTGAGCTTGGCGAATGACAGCCGACCGACCCCGGTACTATGTCGAGCCGC 262  
QY 248 AGCCGAAACCCGAGCCGAGCCGAGCCGAGCCGAAACCGGAAACCGGAACCGG 307  
DB 263 AGCCGAAACCCGAGCCGAGCCGAGCCGAGCCGAAACCGGAAACCGGAACCGG 322  
QY 308 AAGTACCGGATCATCAGAAAGATGATGTTAGGCGGCGCGAGAAATTCATATGATC 367  
DB 323 AAGTACCGGATCATCAGAAAGATGATGTTAGGCGGCGCGAGAAATTCATATGATC 382  
QY 368 TCGAG 372  
DB 383 TCGAG 387  
RESULT 7

AR121624	AR121624	387 bp	DNA	linear	PAT 16-MAY-2001
LOCUS					
DEFINITION	Sequence 16 from patent US 6160088.				
ACCESSION	AR121624				
VERSION	AR121624.1	GI:14105200			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 387)				
TITLE	Rochman,J.E., Mayhew,M. and Hoe,M.H.				
JOURNAL	KDEL receptor inhibitors				
FEATURES	Patent: US 6160088-A 16-12-DEC-2000;				
	Location/Qualifiers				
Source	1..387				
	/organism="unknown"				
BASE COUNT	93 a 107 c 126 g 61 t				
ORIGIN					

Query Match	51.6%	Score 192;	DB 6;	Length 387;
Best Local Similarity	76.4%;	Pred. No. 6.2e-30;		
Matches 307;	Conservative	0;	Mismatches 50;	Indels 45;
				Gaps 4;

QY	1	AAGCTTACCAATGGGAAGGTATCATATTTTAAAGCTTCTGCGCCCTTGGGGCAGTCTGAGC	60
Db	1	AAGCTTACCAATGGGAAGGTATCATATTTTAAAGCTTCTGCGCCCTTGGGGCAGTCTGAGC	59
QY	61	GCTCCCAAAAAAGGATCCAGCCCTGGGTGGAGACTGTTGTCAGACTTGGGCCCCAGATG	120
Db	60	GGGGGCGGAGGATCCAGCCCTGGGTGGAGACTGTTGTCAGACTTGGGCCCCAGATG	105
QY	121	CTTGGGGAACTGCGAGAAACCAACGCGGCGCTGCAGACGTTGCGGGACTGTGCTCGCAG	180
Db	106	CTTGGAGAACTCCAGGAGACTTAAATGCGCGCTGCAGACGTTGAGAGAGACTTTCGACAG	165
QY	181	CAGGTCAGGGGAGATCAAGCTTCTCTAAAAACAAGGATGAGAGTGTGACGCGTGGG----	235
Db	166	CAGGTCAGGGGAGATCACTTCTCTGAAGATACGGGATGAGATGTGACGCTTGGGAGATG	225
QY	236	-----GGCCGAGCGCGACGCGAACCAGCACCCGACGCG	270
Db	226	CAGCCCGACGCGACCCCGCGTACTAAGTTCGACGCGACCGCGAACCAGCACCCGACGCG	285
QY	271	CAGCCGCGACCGGAAACCGCAGCCGGAACCGGAACCGGAAGGTACCGATCATCAGAAAA	330
Db	286	CAGCCGCGACCGGAAACCGCAGCCGGAACCGGAACCGGAAGGTACCGATCATCAGAAAA	345
QY	331	GATGAGTTGTATGGCGCGCGCGAGAAATTCATATGCATCTCGAG	372
Db	346	GATGAGTTGTATGGCGCGCGCGAGAAATTCATATGCATCTCGAG	387

LOCUS	AR121632	444 bp	DNA	linear	PAT 16-MAY-2001
DEFINITION	Sequence 35 from patent US 6160088.				
ACCESSION	AR121632				
VERSION	AR121632.1	GI:14105208			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified:				
AUTHORS	1 (bases 1 to 444)				
TITLE	Rochman,U.E., Mayhew,M. and Hoe,M.H.				
JOURNAL	KDEL receptor inhibitors				
FEATURES	Patent: US 6160088-A 35 12-DEC-2000;				
SOURCE	Location/Qualifiers				
	1..444				
BASE COUNT	115 a	122 c	134 g	73 t	
ORIGIN	/organism="Unknown"				
Query Match	51.3%	Score 190.8;	DB 6;	Length 444;	

[illegible]

RESULT 9					
AR121625					
LOCUS	AR121625	357 bp	DNA	linear	PAT 16-MAY-2001
DEFINITION	Sequence 18 from patent US 6160088.				
ACCESSION	AR121625				
VERSION	AR121625.1	GI:14105201			
KEYWORDS	.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1(bases 1 to 357)				
TITLE	Rothman,J.E., Mayhew,M. and Hoe,M.H.				
JOURNAL	KDEL receptor inhibitors				
FEATURES	Patent: US 6160088-A 18 12-DEC-2000;				
source	Location/Qualifiers				
	1..357				
	/organism="unknown"				
BASE COUNT	84 a 100 c 114 g 59 t				
ORIGIN					

	Query Match	47.7%	Score 177.6	DB 6	Length 357
	Best Local Similarity	73.4%	Pred. No. 6.2e-27		
	Matches 273	Conservative 0	Mismatches 84	Indels 15	Gaps 3
QY	1 AAGCTTACATGGGAAAGTACATGATTTTAAAGCTTGCTCGCCCTTGCGGACATCTGCAGC	60			
DB	1 AAGTTTACATGGGAAAGTTCATCTGTGTGTGGGGGGCTTGCTCTGTGTGGGCGCGGT	59			
QY	61 GCTGCCAAAAAGAGTCCAGCCTGGGTGAGACTGTGTTCAGACCTGGGCGCGCAGATG	120			
DB	60 GCGGGC--CGAGGATCCAGCTGGGTGGAGACTGTGTTAAGCATTTGGTCAACGAG---	114			
QY	121 CTTTCGGGAATCGACGAAACCAACGCGGCGCTGCAGAGACGTGCGGAACTGTGCTCCGACG	180			
DB	115 -----CTCACCCCTTTCAACCAAGATCTTGTGTGAGAGCTTCGGGACGACATCCAGAC	165			
QY	181 CAGGTCAGGGAGATCAAGTTCCTGAAAAACCGGTATGAGAGTGTACGCGTGGCGGCGG	240			
DB	166 CAGGTGAAGAAATGTACTCATCTCCGAAACCATCATATGAGAGTGTCAAGGTGTGCGGTCCG	225			
QY	241 CAGCGCAGCGCGAAACCGCAGCCGACGCGACGCCGACGCGAAACCGCAGCGCGAAACCG	300			
DB	226 CAGCGCAGCGCGAAACCGCAGCCGACGCGACGCCGACGCGAAACCGCAGCGCGAAACCG	285			
QY	301 GAACCGGAAGTACCGGATCATCAGAAAAAGATGAGTGTGTAGCGGCGCGCAGATTCCAT	360			

```

Db      286  GAACCGGAAGGTACCGGATCATCAGAAAAAGATGTTAGCGCGCGGAGAAATTCAT 345
Oy      361  ATGCATCTCGAG 372
Db      346  ATGCATCTCGAG 357

RESULT 10
LOCUS   AR121629          315 bp      DNA      linear      PAT 16-MAY-2001
DEFINITION
Sequence 26 from patent US 6160088.
ACCESSION   AR121629
VERSION     AR121629.1  GI:14105205
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 315)
AUTHORS    Rothman,J.E., Mayhew,M. and Hoe,M.H.
TITLE      KDEL receptor inhibitors
JOURNAL    Patent: US 6160088-A 26 12-DEC-2000;
FEATURES
     source             1..315
                        /organism="unknown"
BASE COUNT      81 a      85 c      79 g      70 t
ORIGIN
Query Match      42.0%; Score 156.4, DB 6; Length 315;
Best Local Similarity 71.0%; Pred. No. 1.6e-22;
Matches 264; Conservative 0; Mismatches 51; Indels 57; Gaps 2;

Oy      1  AAGCTTACATGGGAAGGATCATGATTTTAAAGCTTCTCCGCCCTGGCGGACATCTGCAGC 60
Db      1  AAGCTTACATGGGAAGGATCATGATTTTAAAGCTTCTCCGCCCTGGCGGACATCTGCAGC 60
Oy      61  GCTGCCAAAAAGATCAGCAGCTGGGTGGAGACTGTGTTGCAGACCTGGGCGCCGAGATG 120
Db      61  GCTGCCAAAAAGATCAGCAGCTGGGTGGAGACTGTGTTGCAGAA----- 104
Oy      121  CTTGGGAATCGACGAAACCAACGCGGCTGACAGACGTGCGGAGCTGCTGCGGAG 180
Db      105 -----GCTACAGAAATCTATTATCAATTTCTGTCTC 135
Oy      181  CAGGTCAAGGAGATCAGTTCTCTGAAAAACACGGTGAATGAGTGAAGCCGCGGCGCG 240
Db      136  ATCTTAATATGCTCTTCTGATCTGTATCATCGTGAATGCTTCTC-----CCG 183
Oy      241  CAGCGCAGCCGAAACCGCAGCGCGCAGCGCAGCGGAAACCGCAGCCGAAACCG 300
Db      184  CAGCGCAGCCGAAACCGCAGCGCGCAGCGCAGCGGAAACCGCAGCCGAAACCG 243
Oy      301  GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATTGTAGCGCGCGCAGAAATTCAT 360
Db      244  GAACCGGAAGGTACCGGATCATCAGAAAAAGATGATTGTAGCGCGCGCAGAAATTCAT 303
Oy      361  ATGCATCTCGAG 372
Db      304  ATGCATCTCGAG 315

RESULT 11
LOCUS   AB086984          2274 bp      mRNA      linear      PRI 07-JAN-2003
DEFINITION
Homo sapiens comp mRNA for cartilage oligomeric matrix protein,
complete cds.
ACCESSION   AB086984
VERSION     AB086984.1  GI:27530065
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

REFERENCE   1
AUTHORS     Hashimoto,Y. and Mori,H.
TITLE       Human comp cDNA with 5 SNIPS
JOURNAL     Unpublished
Oy          2 (bases 1 to 2274)
AUTHORS     Hashimoto,Y. and Mori,H.
TITLE       Direct Submission
JOURNAL     Submitted (27-JUN-2002) Hiroshi Mori, Osaka City University Medical
            School, Neuroscience; 1-4-3 Asahimachi, Abeno-ku, Osaka 545-8585,
            Japan (E-mail: moriemed.osaka-cu.ac.jp, Tel:81-6-6645-3920,
            Fax:81-6-6645-3922)
FEATURES
     source             1..2274
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /cell_type="fibroblast"
                        1..2274
                        /gene="comp"
                        1..2274
                        /gene="comp"
                        /codon_start=1
                        /product="cartilage oligomeric matrix protein"
                        /protein_id="BAC53888.1"
                        /db_xref="GI:27530066"
                        /translation="MVPDTACVLLLTALAGASGGGSPGSDIGPQMLREJORTNAA
                        LODVRDMLROOVRREITFLKNTVMEDACAGMOQSVRTGLPSVRPLLHGAPGCFPGVAC
                        IQTSGGRCGCPNGFTNGSHCTDVNENAHNCFPRVRCINTSRGRCACFPQYSG
                        PTHGGVGLAEKAKQVCTDINBCTGQHNKCVPSVCINTGSRFQCGPCQGFVGDQ
                        SGCGRGARPCDPSPECHHADCVLERDGRSVCVRVAGAGNGLICGRPTDLDGFP
                        DEKRCPEPCQKNDKCVTPNSGQEDVDGIDGACPDADGDEVPNEKNDKPLVRNP
                        DORNTDDEKMGDADCNCRSOKNDQKDTDDGDRGDADDIDGRIINQADNCPVRNP
                        SDQSDSGDGTGACDNCNCPKSNPDQADVHDHDFYGDACDSDQDQDGGHDSRNCPT
                        VPSAQRSDSHDGGDADDDDDNDGTPDSDNRRLVYNPQGEADNDGVDGVDQDDE
                        DADRVVKKIDVCPENAEVTLTDFRAFTVVLDPFGDAQIDENWVLLNQREIVTQWS
                        DPGLAVGYTLFNGVDPEGTFRVNTVTDVDAAGFLFGQDSSFYVMMKQMEQYQWA
                        NPFAVAPEPTGLQAVKSTGPGQLRNALHTGDTESQVRLMKDKDRNVMGKDKSY
                        RMFLQHRPQVGYIRVRYEGEELVADSNVVLDTTRMGRLGVCFQSEENIIMANLYR
                        CNDRIIPDYERHQLRQA"
                        150
                        /gene="comp"
                        /note="SNIPS of # 1, 2, 3 are localized at the pentamer
                        formation domain"
                        /replace="g"
                        151
                        /gene="comp"
                        /note="SNIPS of # 1, 2, 3 are localized at the pentamer
                        formation domain"
                        /replace="c"
                        152
                        /gene="comp"
                        /note="SNIPS of # 1, 2, 3 are localized at the pentamer
                        formation domain"
                        /replace="t"
                        326
                        /gene="comp"
                        /note="SNIPS of # 4, 5, 6, 7, 8, 9 are localized at the
                        EGF-like repeats"
                        /replace="c"
                        327
                        /gene="comp"
                        /note="SNIPS of # 4, 5, 6, 7, 8, 9 are localized at the
                        EGF-like repeats"
                        /replace="g"
                        669
                        /gene="comp"
                        /note="SNIPS of # 4, 5, 6, 7, 8, 9 are localized at the
                        EGF-like repeats"
                        /replace="g"
                        670
                        /gene="comp"
                        /note="SNIPS of # 4, 5, 6, 7, 8, 9 are localized at the
                        EGF-like repeats"

```

```
variation /replace="c"
766 /gene="comp"
/notes="SNIPs of # 4, 5, 6, 7, 8, 9 are localized at the
EGF-like repeats"
/replace="g"
767 /gene="comp"
/notes="SNIPs of # 4, 5, 6, 7, 8, 9 are localized at the
EGF-like repeats"
/replace="c"
954 /gene="comp"
/notes="SNIPs of # 10, 11 are localized in a stuffed
sequence between the EGF-like repeats and the TSP types
repeats"
/replace="g"
955 /gene="comp"
/notes="SNIPs of # 10, 11 are localized in a stuffed
sequence between the EGF-like repeats and the TSP types
repeats"
/replace="c"

BASE COUNT 467 a 707 c 749 g 351 t
ORIGIN
Query Match 38.3%; Score 142.6; DB 9; Length 2274;
Best Local Similarity 97.3%; Pred. No. 1e-19;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 157
DB 80 GCTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 139
QY 158 ACGTCCGGGACTGGCTGCGCGCAGCAGTCAAGGAGATCACTTCCTGAAAAACACGGTGA 217
DB 140 ACGTCCGGGACTGGCTGCGCGCAGCAGTCAAGGAGATCACTTCCTGAAAAACACGGTGA 199
QY 218 TGGAGTGTGACCGCTGCGCGCGCCGACGCCG 246
DB 200 TGGAGTGTGACCGCTGCGCGCGATGCACGACG 228

RESULT 12
AX329932 2439 bp DNA linear PAT 09-JAN-2002
LOCUS AX329932
DEFINITION Sequence 441 from Patent WO0194629.
ACCESSION AX329932
VERSION AX329932.1 GI:18102910
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrihan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 441 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..2439
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 503 a 758 c 809 g 369 t
ORIGIN
Query Match 38.3%; Score 142.6; DB 6; Length 2439;
Best Local Similarity 97.3%; Pred. No. 9.9e-20;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 98 GTTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 157
DB 105 GCTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 164
QY 158 ACGTCCGGGACTGGCTGCGCGCAGCAGTCAAGGAGATCACTTCCTGAAAAACACGGTGA 217
DB 165 ACGTCCGGGACTGGCTGCGCGCAGCAGTCAAGGAGATCACTTCCTGAAAAACACGGTGA 224
QY 218 TGGAGTGTGACCGCTGCGCGCGCCGACGCCG 246
DB 225 TGGAGTGTGACCGCTGCGCGCGATGCACGACG 253

RESULT 13
AX330698 2439 bp DNA linear PAT 09-JAN-2002
LOCUS AX330698
DEFINITION Sequence 1207 from Patent WO0194629.
ACCESSION AX330698
VERSION AX330698.1 GI:18103676
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrihan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 1207 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
1..2439
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 503 a 758 c 809 g 369 t
ORIGIN
Query Match 38.3%; Score 142.6; DB 6; Length 2439;
Best Local Similarity 97.3%; Pred. No. 9.9e-20;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 157
DB 105 GCTCAGACCTGGGCCCCGAGATGCTTCGGGAACTGCAGAAACCAACCGCGCGCTGCAGG 164
QY 158 ACGTCCGGGACTGGCTGCGCGCAGCAGTCAAGGAGATCACTTCCTGAAAAACACGGTGA 217
DB 165 ACGTCCGGGACTGGCTGCGCGCAGCAGTCAAGGAGATCACTTCCTGAAAAACACGGTGA 224
QY 218 TGGAGTGTGACCGCTGCGCGCGCCGACGCCG 246
DB 225 TGGAGTGTGACCGCTGCGCGCGATGCACGACG 253

RESULT 14
AX330917 2439 bp DNA linear PAT 09-JAN-2002
LOCUS AX330917
DEFINITION Sequence 1426 from Patent WO0194629.
ACCESSION AX330917
VERSION AX330917.1 GI:18103896
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrihan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 1426 13-DEC-2001;
```

Avalon Pharmaceuticals (US)					
FEATURES					
source location/Qualifiers					
	1..2439	/organism="Homo sapiens"			
		/mol_type="genomic DNA"			
		/db_xref="taxon:9606"			
BASE COUNT	503 a	_758 c	809 g	369 t	
ORIGIN					
Query Match	38.3%	Score 142.6;	DB 6;	Length 2439;	
Best Local Similarity	97.3%;	Pred. No. 9.9e-20;			
Matches 145;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;	
OY	98	GTTGACAGCTGGGCCCGCAGATGCTTGGGGAATCTGCAGAAACCACGCCGGCGCTGCAGG	157		
Db	105	GCTCAGACTTGCCGCCGAGATGCTTCGGGAACTGCAGAAAACCAACCGCGCTGCAGG	164		
OY	158	ACGTGCGGAGCTGCGTCGCGCAGCAGTGCAGGAGATCACGTTCTCTGAAAAACAGGTGA	217		
Db	165	ACGTGCGGAGCTGCGTCGCGCAGCAGTGCAGGAGATCACGTTCTCTGAAAAACAGGTGA	224		
OY	218	TGAGGTGTGACGCGTGTGCGGGCCGACCG	246.		
Db	225	TGAGGTGTGACGCGTGTGCGGGCCGACCG	253		
RESULT 15					
HUMCOMP	2439 bp	mRNA	linear	PRI 15-DEC-1994	
LOCUS	Human germ-line oligomeric matrix protein (COMP) mRNA, complete cds.				
DEFINITION	L32137				
VERSION	L32137.1	GI:602449			
KEYWORDS	germline; matrix protein.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Newton,G., Werenowicz,S., Morton,C.C., Copeland,N.G., Gilbert,D.J., Jenkins,N.A. and Lawler,J.				
TITLE	Characterization of human and mouse cartilage oligomeric matrix protein				
JOURNAL	Unpublished (1994)				
COMMENT	Original source text: Homo sapiens cartilage cDNA to mRNA.				
FEATURES	Location/Qualifiers				
source	1..2439	/organism="Homo sapiens"			
		/mol_type="mRNA"			
		/db_xref="taxon:9606"			
		/map="19p13.1"			
		/cell_type="chondrocyte"			
		/tissue type="cartilage"			
	germline				
	1..2439				
		/gene="COMP"			
	26..2299				
		/gene="COMP"			
		/standard name="cartilage oligomeric matrix protein"			
		/note="putative"			
		/codon_start=1			
		/product="matrix protein"			
		/protein_id="AA57253.1"			
		/db_xref="gi:602450"			
		/translation="MVPTACVLLTLTAALGASGGSGPLGSDIGPMLARELQSTNAALGVDRITLIKNTVMWCDACMGOSVFTGLPSRILHCAKPGCFPVNACIQTSRGRCPCRPAGFTNGSHCTDVNECANHPRRPRCLINTSPGRFCEACPFGYVAGSFCGRGAFAKANKQVCETDINECEGTQHNCVPNSVCINTGSFQCQCGQGFVGDAASGCGRGARFCPPDSPEBCHECHADCVLERDSRCYRWAGNGIICGRTDIDGFEDEKRICEBPQRKDNCVTVPNSGEDVDRIIGACPDADGDGVNEKDCPIVRNPNDORTDEBKMGDACDNCRSGKNDDOKIDDOGRGACDDDIIDGRIRNQADNCRVNPSPDKDGSDGIDGACDNCPKSNPDNVHDIVFDGADACSPDDDGCHODSRNCPVPSNAOEDSHDGGDACDDDDNDUGCPDPSDNKRLVFNPCQEDADRVDGVCDDDFDADRVNDIDVCPENAEVTLTIDFAFQTIVLDPEGDAQIDNNMVLNQGRI VDTNMS			
gene					
CDS					

```

                                DPLGAGVYTAFNQVDFEGTGHVNTVTDVYAGFLFGYQDSSFLYVWMKQMEQTVYQD
                                NPPRAVAABEIGIQAAYKSSSTGPGEQRLNALMHHGDTSESQYRLILMKPRANTGMQDKSY
                                RWFLOQRPOVGIIRVRFYBESPELVAQSNVLDITTMGRGLGVFCFSEONTIIMANLRYR
                                CNNTIPEDYEYTHQLRQA"
sig_peptide
26..85
/gene="COMP"
repeat_region
290..892
/note="putative"
/rpt_family="chromospondin type 2"
/rpt_type=tandem
893..1577
/note="putative"
/rpt_family="chromospondin type 3"
/rpt_type=tandem
2420..2425
/gene="COMP"
polya_signal
2439
/gene="COMP"
polya_site
2439
/gene="COMP"
BASE COUNT      503 a      758 c      809 g      369 t
ORIGIN

Query Match      38.3%; Score 142.6; DB 9; Length 2439;
Best Local Similarity 97.3%; Pred. No. 9.9e-20;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      98      GTTCAGACCTGGGCGCCGACAGTCTTGGGAACTGCAGAGAAACCAACGCGCGCTGCAGG      157
Db      105      GCTCAGACCTGGGCGCCGCGCAATCTTGGGAACTGCAGAGAAACCAACGCGCGCTGCAGG      164
QY      158      ACGTGGCGGAGCTGGCTTCGCGCAGCAGAGTGCAGGAGATCACGTTCTGAAAAACAACGGTGA      217
Db      165      ACGTGGCGGAGCTGGCTTCGCGCAGCAGAGTGCAGGAGATCACGTTCTGAAAAACAACGGTGA      224
QY      218      TGGAGTGTGACGCGTGGCGGCGCCGACGCCG      246
Db      225      TGGAGTGTGACGCGTGGCGGAGATGCAGCAG      253

```

11th grade Blank (usps) (order)



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2003, 15:10:25 ; Search time 207 Seconds

(without alignments)  
4851.164 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 372  
Sequence: 1 aagcttaccatgggaagta.....attccatcgcctcgcag 372

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	
1	357	96.0	369	21	AAZ50497	
2	256.2	68.9	21	AAZ50500	KOEL receptor inh1	
3	249.8	67.2	369	21	AAZ50499	KOEL receptor inh1
4	202	54.3	369	21	AAZ50496	KOEL receptor inh1
5	194	52.2	369	21	AAZ50495	KOEL receptor inh1
6	193.4	52.0	387	21	AAZ50492	KOEL receptor inh1
7	192	51.6	387	21	AAZ50493	KOEL receptor inh1
8	190.8	51.3	444	21	AAZ50501	KOEL receptor inh1

9	177.6	47.7	357	21	AAZ50494
10	142.6	38.3	755	21	AAA47734
11	142.6	38.3	925	21	AAA47735
12	142.6	38.3	2439	24	ABT10887
13	142.6	38.3	2439	24	ABT07751
14	142.6	38.3	2439	24	ABL62104
15	142.6	38.3	2439	24	ABL62870
16	142.6	38.3	2439	24	ABL63089
17	142.6	38.3	2439	25	ACC50112
18	142.6	38.3	2439	25	ABX76334
19	141.4	38.0	312	21	AAZ50498
20	137.8	37.0	764	22	AAK91566
21	137.8	37.0	764	22	AAK91336
22	137.8	37.0	1779	22	AAK94103
23	83.6	22.5	8532	22	ABA15407
24	83.6	22.5	8532	22	ABA15585
25	83.6	22.5	46275	24	ABT10145
26	73.8	19.8	78	19	AAV27572
27	58	15.6	34427	20	AAK07371
28	58	15.6	34616	24	ABK49009
29	58	15.6	35937	24	ABK69881
30	58	15.6	36335	15	AAK68003
31	56.4	15.2	480	17	AAK41508
32	55	14.8	1037	21	AAA59242
33	55	14.8	1159	21	AAA59240
34	55	14.8	1472	21	AAA59241
35	53.8	14.5	397	20	AAK69891
36	53.6	14.4	403	21	ABK10240
37	53.6	14.4	5120	22	AAK64677
38	53.4	14.4	1865	23	ABL03153
39	53.4	14.4	5178	23	ABL03152
40	53	14.2	521	21	AAK66020
41	53	14.2	521	24	ABQ92425
42	53	14.2	521	24	ABL49235
43	53	14.2	1082	21	AAK78047
44	53	14.2	1412	24	ABK4787
45	53	14.2	3747	25	ACA03912

## ALIGNMENTS

RESULT 1  
AAZ50497  
ID AAZ50497 standard; DNA; 369 BP.

XX AAZ50497;

AC 23-MAY-2000 (first entry)

XX KDEL receptor inhibitor-6 DNA.

XX KDEL receptor inhibitor; heat shock protein; immune response;

KW oligomerization domain; neoplasia; sarcoma; lymphoma; leukaemia;

KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;

KW infectious disease; allergy; autoimmune disease; ss.

XX Chimeric - Adenovirus E3.

OS Chimeric - Homo sapiens.

OS Chimeric - Camelus sp.

XX Key

FT CDS

FT sig\_peptide

FT mat\_peptide

FT misc\_feature

KDEL receptor inh  
Human COMP/TSP-1 c  
Human COMP/TSP-2 c  
Human breast cance  
Breast cancer-asso  
Colon adenocarcino  
Breast cancer rela  
Breast cancer rela  
Breast cancer asso  
Lung cancer-associ  
KDEL receptor inh  
Human cDNA 5'-end  
Human cDNA clone x  
Human full-length  
Human nervous syst  
Human nervous syst  
Human breast cance  
Nucleotide sequenc  
Adenovirus vector  
Adenoviral vector  
Human adeno-associ  
Ad2/-ORF6/PCP-CFTR  
Human adenovirus t  
Exons E, C and A o  
An EcoRI fragment  
Exons D, C, B and  
Spincerebellar at  
Trinucleotide repe  
DNA sequence of hu  
Drosophila melanog  
Drosophila melanog  
Human lung cancer-  
Human lung tumor  
Human lung cancer  
Human cancer assoc  
Human cDNA differe  
cDNA downregulated

```

FT      misc_feature /note="Human COMP pentamerisation domain"
FT      235..306 /*tag= e
FT      /note="Camel IgG linker domain"
XX
XX      WO200006729-A1.
XX
XX      10-FEB-2000.
XX
XX      28-JUL-1999; 99WO-US17147.
XX
XX      29-JUL-1998; 98US-0124671.
XX
XX      (SLOK ) SLOAN KETTERING INST CANCER RES.
XX      Rothman JE, Mayhew M, Hoe MH;
XX      WPI: 2000-195296/17.
XX      P-PSDB; AAY44963.
XX
XX      Inhibitors of the KDEL receptor which comprises an oligomerization
XX      PT domain useful for promoting secretion of proteins which are normally
XX      PT retained within the cell -
XX
XX      Disclosure; Fig 6; 87pp; English.
XX
XX      The patent discloses the use of KDEL receptor inhibitor to promote
XX      CC secretion of proteins that are normally retained within the cell such as
XX      CC heat shock proteins by inhibiting KDEL receptor-mediated return of
XX      CC protein complexes to endoplasmic reticulum. This makes the secreted heat
XX      CC shock proteins more accessible to the immune system and improves immune
XX      CC response to a target antigen. The inhibitor protein comprises several
XX      CC subunits where each subunit comprises an oligomerisation domain and has
XX      CC at its carboxy terminus a region which binds to a KDEL receptor. The
XX      CC target antigen may be associated with diseases including neoplasia such
XX      CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
XX      CC astrocytoma, with defective tumour suppressor genes, oncogenes,
XX      CC infectious diseases, allergy or autoimmune diseases. The present
XX      CC sequence encodes KDEL receptor inhibitor comprising regions encoding a
XX      CC cleavable signal peptide; the oligomerisation domain of human cartilage
XX      CC oligomeric matrix protein (COMP) pentamerisation domain; a camel IgG
XX      CC linker domain and the carboxy-terminal sequence KDEL. The subsequence
XX      CC GDCC is an alteration of rat COMP which provides increased stability via
XX      CC disulphide bonds.
XX
XX      Sequence 369 BP; 90 A; 102 C; 120 G; 57 T; 0 other;
XX
XX      Query Match 96.0%; Score 357; DB 21; Length 369;
XX      Best Local Similarity 100.0%; Pred. No. 6.2e-76;
XX      Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      16 AGGTACATGATTTTGGCTTGCCTGCGCGGAGTGTGAGCGGCTGCCAAAAAAGA 75
XX      |||||
XX      13 AGGTACATGATTTTGGCTTGCCTGCGCGGAGTGTGAGCGGCTGCCAAAAAAGA 72
XX
XX      76 TCCAGCTGGGTGAGACCTGTTTTCAGACCTGGGCGCGCAGATGCTTGGGAACTGCAG 135
XX      |||||
XX      73 TCCAGCTGGGTGAGACCTGTTTTCAGACCTGGGCGCGCAGATGCTTGGGAACTGCAG 132
XX
XX      136 GAAACCAACGCGCGCTGCGAGACCTGCGGAGTGTGAGCGGCTGCCAGAGTCAAGAGATC 195
XX      |||||
XX      133 GAAACCAACGCGCGCTGCGAGACCTGCGGAGTGTGAGCGGCTGCCAGAGTCAAGAGATC 192
XX
XX      196 AGCTTCTGTAACCAACGCGTGTGAGTGTGAGCGGCTGCCAGAGTCAAGAGATC 255
XX      |||||
XX      193 AGCTTCTGTAACCAACGCGTGTGAGTGTGAGCGGCTGCCAGAGTCAAGAGATC 252
XX
XX      256 CCGCAGCCGCGAGCCGCGAGCCGGAACCGCAGCCGGAACCGGAACCGGAAGGTACC 315
XX      |||||
XX      253 CCGCAGCCGCGAGCCGCGAGCCGGAACCGCAGCCGGAACCGGAACCGGAAGGTACC 312
XX
XX      316 GGATCATCGAAAAAGATGATTTGTTAGGGCGCGCAGAAATTCATATGATCTCCAG 372
XX      |||||

```

```

DB      313 GGATCATCGAAAAAGATGATTTGAGCGCGCGCAGAAATTCATATGATCTCCAG 369
XX
XX      RESULT 2
XX      ID AAZ50500 standard; DNA; 369 BP.
XX
XX      AAZ50500;
XX
XX      23-MAY-2000 (first entry)
XX
XX      KDEL receptor inhibitor-9 DNA.
XX
XX      KDEL receptor inhibitor; heat shock protein; immune response;
XX      KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
XX      KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX      KW infectious disease; allergy; autoimmune disease; ss.
XX
XX      Chimeric - Adenovirus E3.
XX      OS Chimeric - Homo sapiens.
XX      OS Chimeric - Camelus sp.
XX
XX      Key Location/Qualifiers
XX      FT CDS 10..339
XX      FT /*tag= a
XX      FT /product= "KDEL receptor inhibitor protein"
XX      FT sig_peptide 10..69
XX      FT /*tag= b
XX      FT /note= "Derived from adenovirus E3"
XX      FT mat_peptide 70..336
XX      FT /*tag= c
XX      FT /note= "Mature KDEL receptor inhibitor protein"
XX      FT misc_feature 97..234
XX      FT /*tag= d
XX      FT /note= "Human TSP4 trimerisation domain"
XX      FT misc_feature 235..306
XX      FT /*tag= e
XX      FT /note= "Camel IgG linker domain"
XX
XX      WO200006729-A1.
XX
XX      10-FEB-2000.
XX
XX      28-JUL-1999; 99WO-US17147.
XX
XX      29-JUL-1998; 98US-0124671.
XX
XX      (SLOK ) SLOAN KETTERING INST CANCER RES.
XX      Rothman JE, Mayhew M, Hoe MH;
XX      WPI: 2000-195296/17.
XX      P-PSDB; AAY44966.
XX
XX      Inhibitors of the KDEL receptor which comprises an oligomerization
XX      PT domain useful for promoting secretion of proteins which are normally
XX      PT retained within the cell -
XX
XX      Disclosure; Fig 9; 87pp; English.
XX
XX      The patent discloses the use of KDEL receptor inhibitor to promote
XX      CC secretion of proteins that are normally retained within the cell such as
XX      CC heat shock proteins by inhibiting KDEL receptor-mediated return of
XX      CC protein complexes to endoplasmic reticulum. This makes the secreted heat
XX      CC shock proteins more accessible to the immune system and improves immune
XX      CC response to a target antigen. The inhibitor protein comprises several
XX      CC subunits where each subunit comprises an oligomerisation domain and has
XX      CC at its carboxy terminus a region which binds to a KDEL receptor. The
XX      CC target antigen may be associated with diseases including neoplasia such
XX      CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
XX      CC astrocytoma, with defective tumour suppressor genes, oncogenes,
XX      CC infectious diseases, allergy or autoimmune diseases. The present
XX      CC sequence encodes KDEL receptor inhibitor comprising regions encoding a

```

cleavable signal peptide; the oligomerisation domain of human thrombospondin 4 (TSP4) trimerisation domain; a camel IgG linker domain and the carboxy-terminal sequence KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric matrix protein which provides increased stability via disulphide bonds.

Sequence 369 BP; 101 A; 96 C; 104 G; 68 T; 0 other;

Query Match 68.9%; Score 256.2; DB 21; Length 369;  
Best Local Similarity 82.4%; Pred. No. 6,8e-52;  
Matches 294; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

16 AGGTACATGATTTTAAAGCTTCTCCCTTGGCGAGCTGCGAGCGTCCAAAAAGAA 75  
13 AGGTACATGATTTTAAAGCTTCTCCCTTGGCGAGCTGCGAGCGTCCAAAAAGAA 72  
76 TCCAGCTGGGTGAGAGCTGTTTTCAGACTGGGCGCGAGAGCTTGGGAATGCA 135  
73 TCCAGCTGGGTGAGAGCTGTTTTCAGACTGGGCGCGAGAGCTTGGGAATGCA 132  
136 GAAACCAAGCGGCGCTGCGAGAGCTGGGCGAGCTGGCGAGAGCTGCGAGATC 195  
133 CAATTAAACCACTCTGAGAGAGGTGAAGAGCTTCTGAGACAGAGTTAAGAAACA 192  
196 AGCTTCTGAAAAACACCGGTGATGAGTGTGACGGGTGCGGCGCGAGCGCGAA 255  
193 TCATTTTTGGCAAAACACATAGCTGAATGCCAGCTTGCGGCTCCAGCCGAGAA 252  
256 CCGAGCGCGAGCGCGAGCGCGAGCGCGAAACCGGAGCGGAAACCGGAAGTACC 315  
253 CCGAGCGCGAGCGCGAGCGCGAGCGCGAAACCGGAGCGGAAACCGGAAGTACC 312  
316 GGATCATCAGAAAAAGATGATGATGAGCGGCGCGAGAGTTCATATGATCTGAG 372  
313 GGATCATCAGAAAAAGATGATGATGAGCGGCGCGAGAGTTCATATGATCTGAG 369

RESULT 3

AAZ50499 standard; DNA; 369 BP.

AAZ50499;

23-MAY-2000 (first entry)

KDEL receptor inhibitor-8 DNA.

KDEL receptor inhibitor; heat shock protein; immune response;

oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;

melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;

infectious disease; allergy; autoimmune disease; se.

Chimeric - Adenovirus E3.

Chimeric - Homo sapiens.

Chimeric - Camelus sp.

Key Location/Qualifiers

CDS 10..339

FT /tag= a

FT /product= "KDEL receptor inhibitor protein"

FT 10..69

FT /tag= b

FT /note= "Derived from adenovirus E3"

FT mat\_peptide 70..336

FT /tag= c

FT /note= "Mature KDEL receptor inhibitor protein"

FT misc\_feature 97..234

FT /tag= d

FT /note= "Human TSP3 trimerisation domain"

FT misc\_feature 235..306

FT /tag= e

FT /note= "Camel IgG linker domain"

XX

PN WO200006729-A1.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17147.

XX 29-JUL-1998; 98US-0124671.

XX (SLOK ) SLOAN KETTERING INST CANCER RES.

XX Rothman JE, Mayhew M, Hoe MH;

XX WPI; 2000-195296/17.

XX P-PSDB; AAY44965.

PT Inhibitors of the KDEL receptor which comprises an oligomerization

PT domain useful for promoting secretion of proteins which are normally

PT retained within the cell

PS Disclosure; Fig 8; 87pp; English.

The patent discloses the use of KDEL receptor inhibitor to promote secretion of proteins that are normally retained within the cell such as heat shock proteins by inhibiting KDEL receptor-mediated return of protein complexes to endoplasmic reticulum. This makes the secreted heat shock proteins more accessible to the immune system and improves immune response to a target antigen. The inhibitor protein comprises several subunits where each subunit comprises an oligomerisation domain and has at its carboxy terminus a region which binds to a KDEL receptor. The target antigen may be associated with diseases including neoplasia such as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and infectious diseases, allergy or autoimmune diseases. The present sequence encodes KDEL receptor inhibitor comprising regions encoding a cleavable signal peptide; the oligomerisation domain of human thrombospondin 3 (TSP3) trimerisation domain; a camel IgG linker domain and the carboxy-terminal sequence KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric matrix protein which provides increased stability via disulphide bonds.

XX Sequence 369 BP; 95 A; 105 C; 107 G; 62 T; 0 other;

Query Match 67.2%; Score 249.8; DB 21; Length 369;

Best Local Similarity 81.2%; Pred. No. 2,3e-50;

Matches 290; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

16 AGGTACATGATTTTAAAGCTTCTCCCTTGGCGAGCTGCGAGCGTCCAAAAAGAA 75  
13 AGGTACATGATTTTAAAGCTTCTCCCTTGGCGAGCTGCGAGCGTCCAAAAAGAA 72  
76 TCCAGCTGGGTGAGAGCTGTTTTCAGACTGGGCGCGAGAGCTTGGGAATGCA 135  
73 TCCAGCTGGGTGAGAGCTGTTTTCAGACTGGGCGCGAGAGCTTGGGAATGCA 132  
136 GAAACCAAGCGGCGCTGCGAGAGCTGGGCGAGCTGGGCGAGAGCTGCGAGATC 195  
133 CTCTTCAACCAAGCTTCTGAGAGCTTGGGCGAGAGCTGCGAGAGCTGCGAGATC 192  
196 AGCTTCTGAAAAACACCGGTGATGAGTGTGACGGGTGCGGCGCGAGCGCGAA 255  
193 TCATCATCGGAAACACCATCATGAGTGTGAGTGTGCGGCTCCAGCCGAGCGAA 252  
256 CCGAGCGCGAGCGCGAGCGCGAGCGCGAAACCGGAGCGGAAACCGGAAGTACC 315  
253 CCGAGCGCGAGCGCGAGCGCGAGCGCGAAACCGGAGCGGAAACCGGAAGTACC 312  
316 GGATCATCAGAAAAAGATGATGATGAGCGGCGCGAGAGTTCATATGATCTGAG 372  
313 GGATCATCAGAAAAAGATGATGATGAGCGGCGCGAGAGTTCATATGATCTGAG 369

RESULT 4

AAZ50496

1D	AA50496 standard; DNA; 369 BP.
XX	
AC	AA50496;
XX	
DT	23-MAY-2000 (first entry)
XX	
DE	KDEL receptor inhibitor-5 DNA.
XX	
KW	KDEL receptor inhibitor; heat shock protein; immune response;
KW	oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
KW	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX	infectious disease; allergy; autoimmune disease; ss.
OS	Chimeric - Mus sp.
OS	Chimeric - Xenopus sp.
OS	Chimeric - Camelus sp.
XX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	10..339
FT	/*tag= a
FT	/product= "KDEL receptor inhibitor protein"
FT	10..68
FT	/*tag= b
FT	/note= "Derived from mouse Bp"
FT	69..336
FT	/*tag= c
FT	/product= "Mature KDEL receptor inhibitor protein"
FT	97..234
FT	/*tag= d
FT	/note= "Xenopus TSP4 trimerisation domain"
FT	235..306
FT	/*tag= e
FT	/note= "Camel IgG linker domain"
XX	
PN	WO200006729-A1.
XX	
PD	10-FEB-2000.
XX	
PP	28-JUL-1999; 99WO-US17147.
XX	
PR	29-JUL-1998; 98US-0124671.
XX	
PA	(SLOK ) SLOAN KETTERING INST CANCER RES.
XX	
P1	Rothman JE, Mayhew M, Hoe MH;
PI	WPI: 2000-195296/17.
DR	P-PSDB; AAY44962.
XX	
PT	Inhibitors of the KDEL receptor which comprises an oligomerization
PT	domain useful for promoting secretion of proteins which are normally
PT	retained within the cell -
XX	
PS	Disclosure; Fig 5; 87pp; English.
XX	
CC	The patent discloses the use of KDEL receptor inhibitor to promote
CC	secretion of proteins that are normally retained within the cell such as
CC	heat shock proteins by inhibiting KDEL receptor-mediated return of
CC	protein complexes to endoplasmic reticulum. This makes the secreted heat
CC	shock proteins more accessible to the immune system and improves immune
CC	response to a target antigen. The inhibitor protein comprises several
CC	subunits where each subunit comprises an oligomerisation domain and has
CC	at its carboxy terminus a region which binds to a KDEL receptor. The
CC	target antigen may be associated with diseases including neoplasia such
CC	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
CC	astrocytoma, with defective tumour suppressor genes, oncogenes,
CC	infectious diseases, allergy or autoimmune diseases. The present
CC	sequence encodes KDEL receptor inhibitor comprising regions encoding a
CC	cleavable signal peptide; the oligomerisation domain of Xenopus
CC	thrombospondin 4 (TSP4) trimerisation domain including an additional
CC	sub-sequence; a camel IgG linker domain and the carboxy-terminal sequence
CC	KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric
CC	matrix protein which provides increased stability via disulphide bonds.
CC	

XX	Sequence	369 BP, 94 A, 94 C, 121 G, 60 T, 0 other;
XX	Query Match	54.3%; Score 202; DB 21; Length 369;
XX	Best Local Similarity	79.1%; Pred. No. 5.8e-39;
XX	Matches 239; Conservative	0; Mismatches 63; Indels 0; Gaps 0;
OY	71 AAGATCCAGCCTTGAGAGACTGTGTTTCAGACCTGGGCCCCGAGATGCTTCGGGAAC	130
DB	68 AGGATCCAGCCTGGGTGAGACTGTGTGGTGACGTACAGACAGATTGATGGCCAGA	127
OY	131 TGCAGGAACCAACGCGCGCTGAGAGACGTCGCGGACTGGCTCGGAGAGGTCAAGG	190
DB	128 TAACCCAAATGATCAGATGCTGGAGAGCTCCAGATGTCATGAGACAGACGTTGAAG	187
OY	191 AGATCAGCTTCTCTGAAAAACACGGTGAATGAGTGTGACGCGTGGCGGCGGACCGCAGC	250
DB	188 AGACCATGTTCTTTAGAAAAACCAATTCGACGAATCCAGGCTGTGTGCCGACCGCAGC	247
OY	251 CGAAACCGCAGCCGACGCGCAGCGCCGACCGGAAACCGCAGCCGAAACCGGAAC	310
DB	248 CGAAACCGCAGCCGACGCGCAGCGCCGACCGGAAACCGCAGCCGAAACCGGAAG	307
OY	311 GTACCGGATCATCGAAAAAAGATGTTGTGTGGGGCGGCGAGAAATTCATATGATCTTCG	370
DB	308 GTACCGGATCATCGAAAAAAGATGTTGTGTGGGGCGGCGAGAAATTCATATGATCTTCG	367
OY	371 AG 372	
DB	368 AG 369	
XX	RESULT 5	
XX	AAZ50495	
XX	AAZ50495 standard; DNA; 369 BP.	
XX	AAZ50495;	
XX	23-MAY-2000 (first entry)	
XX	KDEL receptor inhibitor-4 DNA.	
XX	KDEL receptor inhibitor; heat shock protein; immune response;	
KW	oligomerization domain; neoplasia; sarcoma; lymphoma; leukaemia;	
KW	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;	
KW	infectious disease; allergy; autoimmune disease; ss.	
XX	Chimeric - Mus sp.	
OS	Chimeric - Camelus sp.	
XX	Key	Location/Qualifiers
FT	CDS	10..339
FT		/*tag= a
FT		/product= "KDEL receptor inhibitor protein"
FT	sig_peptide	10..68
FT		/*tag= b
FT		/note= "Derived from mouse Bip"
FT	mat_peptide	69..336
FT		/*tag= C
FT		/product= "Mature KDEL receptor inhibitor protein"
FT	misc_feature	97..234
FT		/*tag= d
FT		/note= "Mouse TSP3 domain"
FT	misc_feature	235..306
FT		/*tag= e
FT		/note= "Camel IgG linker domain"
XX	WO200006729-A1.	
XX	10-FEB-2000.	
XX	28-JUL-1999; 99WO-US17147.	

```

PR 29-JUL-1998; 98US-0124671.
PA (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
XX
XX Rothman JE, Mayhew M, Hoe MH;
XX WPI; 2000-195296/17.
XX P-PSDB; AAY44961.
XX
XX Inhibitors of the KDEL receptor which comprises an oligomerization
XX domain useful for promoting secretion of proteins which are normally
XX retained within the cell -
XX
XX Disclosure; Fig 4; 87pp; English.
XX
XX The patent discloses the use of KDEL receptor inhibitor to promote
XX secretion of proteins that are normally retained within the cell such as
XX heat shock proteins by inhibiting KDEL receptor-mediated return of
XX protein complexes to endoplasmic reticulum. This makes the secreted heat
XX shock proteins more accessible to the immune system and improves immune
XX response to a target antigen. The inhibitor protein comprises several
XX subunits where each subunit comprises an oligomerisation domain and has
XX at its carboxy terminus a region which binds to a KDEL receptor. The
XX target antigen may be associated with diseases including neoplasia such
XX as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
XX astrocytoma, with defective tumour suppressor genes, oncogenes,
XX infectious diseases, allergy or autoimmune diseases. The present
XX sequence encodes KDEL receptor inhibitor comprising regions encoding a
XX cleavable signal peptide; the oligomerisation domain of mouse
XX thrombospondin 3 (TSP3) trimerisation domain including an additional
XX sub-sequence; a camel IgG linker domain and the carboxy-terminal sequence
XX KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric
XX matrix protein which provides increased stability via disulphide bonds.
XX
XX Sequence 369 BP; 87 A; 103 C; 120 G; 59 T; 0 other;
XX
XX Query Match 52.2%; Score 194; DB 21; Length 369;
XX Best Local Similarity 77.5%; Pred. No. 4,7e-37;
XX Matches 234; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
XX
XX 71 AAGATCCAGCTTGGTGAGAGCTTTGTCAGACCTTGGGCGCGAGATGCTTCGGGAAC 130
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 68 AGGATCCAGCTTGGTGAGAGCTTTGTCAGACCTTGGGCGCGAGATGCTTCGGGAAC 127
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 131 TGCAGGAACCAACCGCGCGCTGCGAGACGTGCGGGAAGTGGCGGAGAGGTCAAGG 190
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 128 TCACCTCTTCAACCAAGATCTTGAAGCTTCCGGGAGCAATCCGAGACCGAGTGAAGG 187
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 191 AGATCAGCTTCTGTAACCAACGATGATGAGTGAACGCGCGCGCGAGCGCGAGC 250
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 188 AATATCTACTATCCGGAACACCATCATGAGTGTGAGTGTGCGGCGCGAGCGCGAGC 247
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 251 CGAAACCGCGAGCGCGAGCGCGAGCGCGAGAACCGGAGCCGGAACCGGAAG 310
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 248 CGAAACCGCGAGCGCGAGCGCGAGCGCGAGAACCGGAGCCGGAACCGGAAG 307
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 311 GTACCGGATCATCAAAAAAGTGTGTAGCGCGCGCGAGAAATTCATATGCACTTCG 370
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 308 GTACCGGATCATCAAAAAAGTGTGTAGCGCGCGCGAGAAATTCATATGCACTTCG 367
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 371 AG 372
XX | |
XX 368 AG 369
XX
XX RESULT 6
XX ID AAZ50492 standard; DNA; 387 BP.
XX
XX AAZ50492;
XX
XX 23-MAY-2000 (first entry)
XX

```

```

DE KDEL receptor inhibitor-1 DNA.
XX
XX KDEL receptor inhibitor; heat shock protein; immune response;
XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX infectious disease; allergy; autoimmune disease; ss.
XX
XX Chimeric - Mus sp.
XX OS Chimeric - Camelus sp.
XX OS Chimeric - Rattus sp.
XX
XX Key Location/Qualifiers
XX CDS 10..357
XX FT /tag= a
XX FT /product= "KDEL receptor inhibitor protein"
XX
XX sig_peptide 10..68
XX FT /tag= b
XX FT /note= "Derived from mouse Bp"
XX
XX mat_peptide 69..354
XX FT /tag= c
XX FT /product= "Mature KDEL receptor inhibitor protein"
XX
XX misc_feature 98..222
XX FT /tag= d
XX FT /note= "Rat COMP domain"
XX
XX misc_feature 253..324
XX FT /tag= e
XX FT /note= "Camel IgG linker domain"
XX
XX
XX WO200006729-A1.
XX
XX 10-FEB-2000.
XX
XX 28-JUL-1999; 99WO-US17147.
XX
XX 29-JUL-1998; 98US-0124671.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
XX Rothman JE, Mayhew M, Hoe MH;
XX WPI; 2000-195296/17.
XX P-PSDB; AAY44958.
XX
XX Inhibitors of the KDEL receptor which comprises an oligomerization
XX domain useful for promoting secretion of proteins which are normally
XX retained within the cell -
XX
XX Disclosure; Fig 1; 87pp; English.
XX
XX The patent discloses the use of KDEL receptor inhibitor to promote
XX secretion of proteins that are normally retained within the cell such as
XX heat shock proteins by inhibiting KDEL receptor-mediated return of
XX protein complexes to endoplasmic reticulum. This makes the secreted heat
XX shock proteins more accessible to the immune system and improves immune
XX response to a target antigen. The inhibitor protein comprises several
XX subunits where each subunit comprises an oligomerisation domain and has
XX at its carboxy terminus a region which binds to a KDEL receptor. The
XX target antigen may be associated with diseases including neoplasia such
XX as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and
XX astrocytoma, with defective tumour suppressor genes, oncogenes,
XX infectious diseases, allergy or autoimmune diseases. The present
XX sequence encodes KDEL receptor inhibitor comprising regions encoding a
XX cleavable signal peptide; the oligomerisation domain from rat cartilage
XX oligomeric matrix protein; a camel IgG linker domain and the carboxy
XX -terminal sequence KDEL. This is introduced into host cells by suitable
XX vectors.
XX
XX Sequence 387 BP; 94 A; 110 C; 125 G; 58 T; 0 other;
XX
XX Query Match 52.0%; Score 193.4; DB 21; Length 387;
XX Best Local Similarity 81.6%; Pred. No. 6,6e-37;
XX Matches 249; Conservative 0; Mismatches 26; Indels 30; Gaps 1;
XX

```

```

QY 98 GTTCAGACCTGGGCCCCGACGATGCTTCGGAACTGCAGAAACCAACCGCGCGCTGCAG 157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 GTGAGACCTAGACCCCAAGATGCTTCGAGAACTCCAGACACTAATGCGCGCTGCAAG 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 158 AGGTGCGGAGCTGGCTGGCGGACGAGTCAGAGATCACTTCGAAAAACAGGTTA 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 AGGTGAGAGAGCTCTTCGACAGACAGGTCAAGAGATCACTTCCTGAAAGATACGTTA 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 218 TCGAGTGTACCGCTGCG-----GGCGGACGCGCG 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 TCGAATGTGACCGTTCGGAATGACAGCCGCAACGCCCGGTAATGTCGCGACGCG 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 248 ACCCGAAACCGGACCGCGGACCGCGACCGCGCAACCGCAACCGGAAACCGG 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 ACCCGAAACCGGACCGCGGACCGCGACCGCGCAACCGCGCAACCGGAAACCGG 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 308 AAGGTACCGGATCATGACAAAGATGATGAGCGGCGCGAGAAATTCATATGATC 367
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 AAGGTACCGGATCATGACAAAGATGATGAGCGGCGCGAGAAATTCATATGATC 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 368 TCGAG 372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 TCGAG 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 7
AAZ50493
ID AAZ50493 standard; DNA; 387 BP.
XX
AC AAZ50493;
XX
DT 23-MAY-2000 (first entry)
DE
DE KDEL receptor inhibitor-2 DNA.
XX

```

```

XX KDEL receptor inhibitor; heat shock protein; immune response;
KM oligomerization domain; neoplasia; sarcoma; lymphoma; leukaemia;
KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX infectious disease; allergy; autoimmune disease; ss.
XX
OS Chimeric - Mus sp.
OS Chimeric - Camelus sp.
OS Chimeric - Rattus sp.
XX

```

```

FH Key Location/Qualifiers
FT CDS 10..357
FT FT /tag= a
FT FT /product= "KDEL receptor inhibitor protein"
FT FT 10..68
FT FT /tag= b
FT FT /note= "Derived from mouse Bip"
FT FT 69..354
FT FT /tag= c
FT FT /product= "Mature KDEL receptor inhibitor protein"
FT FT 98..222
FT FT /tag= d
FT FT /note= "Rat COMP domain"
FT FT 253..324
FT FT /tag= e
FT FT /note= "Camel IgG linker domain"
XX
XX WO200006729-A1.
XX
XX 10-FEB-2000.
XX
XX 28-JUL-1999; 99WO-US17147.
XX
XX 29-JUL-1998; 98US-0124671.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
XX Rothman JE, Mayhew M, Hoe MH;
XX
XX

```

```

DR WPI; 2000-195296/17.
DR P-PSDB; AAY44959.
XX
PT Inhibitors of the KDEL receptor which comprises an oligomerization
PT domain useful for promoting secretion of proteins which are normally
PT retained within the cell -
XX
PS Disclosure; Fig 2; 87pp; English.
XX

```

The patent discloses the use of KDEL receptor inhibitor to promote secretion of proteins that are normally retained within the cell such as heat shock proteins by inhibiting KDEL receptor-mediated return of protein complexes to endoplasmic reticulum. This makes the secreted heat shock proteins more accessible to the immune system and improves immune response to a target antigen. The inhibitor protein comprises several CC subunits where each subunit comprises an oligomerization domain and has at its carboxy terminus a region which binds to a KDEL receptor. The target antigen may be associated with diseases including neoplasia such as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and astrocytoma, with defective tumour suppressor genes, oncogenes, infectious diseases, allergy or autoimmune diseases. The present sequence encodes KDEL receptor inhibitor comprising regions encoding a CC cleavable signal peptide; the oligomerization domain from rat cartilage oligomeric matrix protein (COMP); a camel IgG linker domain and the CC carboxy-terminal sequence KDEL. The subsequence GGCC is an alteration of CC rat COMP which provides increased stability via disulphide bonds. This is introduced into host cells by suitable vectors.

Sequence 387 BP; 93 A; 107 C; 126 G; 61 T; 0 other;

Query Match 51.6%; Score 192; DB 21; Length 387;

Best Local Similarity 76.4%; Pred. No. 1.4e-36;

Matches 307; Conservative 0; Mismatches 50; Indels 45; Gaps 4;

```

QY 1 AAGCTTACATGAGGAGGATGATGATTTAGGCTTGTGCGCTTGGCGGAGTGTGACG 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 AAGCTTACATGAGGAGGAGGATGATGATTTAGGCTTGTGCGCTTGGCGGAGTGTGACG 59
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 GTGCGCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 GCGGCG--CGAGGATTCAGGCTGTGATGATGATGATGATGATGATGATGATGATGATG 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 CTTGCGGAACTGCGAGAAACCAACCGCGCTGCGAGACGTCGCGGACTGCTGCGGAG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 CTTGCGGAACTGCGAGAAACCAACCGCGCTGCGAGACGTCGCGGACTGCTGCGGAG 165
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 181 CAGGTCAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 CAGGTCAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 236 -----GGCGGACCGGACCGGAAACCGGAGTGTGAGCGCGG----- 270
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 CAGCGCGACGACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 271 CAGCGCGACGAGAAACCGGAGCGGAAACCGGAGTGTGATGATGATGATGATGATG 330
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 CAGCGCGACGAGAAACCGGAGCGGAAACCGGAGTGTGATGATGATGATGATGATG 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 331 GATGAGTGTGAGGCGGCGGAGATGATGATGATGATGATGATGATGATGATGATGATG 372
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 GATGAGTGTGAGGCGGCGGAGATGATGATGATGATGATGATGATGATGATGATGATG 387
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 8
AAZ50501
ID AAZ50501 standard; DNA; 444 BP.
XX
AC AAZ50501;
XX
DT 23-MAY-2000 (first entry)
XX
DE KDEL receptor inhibitor-10 DNA.
XX
XX

```

KM KDEL receptor inhibitor; heat shock protein; immune response;  
 KM oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;  
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
 KM infectious disease; allergy; autoimmune disease; ss.  
 OS Chimeric - Mus sp.  
 OS Chimeric - Rattus sp.  
 OS Chimeric - Camelus sp.  
 XX Key Location/Qualifiers  
 FT CDS 10..414  
 FT /tag= a  
 FT /product= "KDEL receptor inhibitor protein"  
 FT sig\_peptide 10..69  
 FT /tag= b  
 FT /note= "Derived from mouse Bip"  
 FT mat\_peptide 70..411  
 FT /tag= c  
 FT /product= "Mature KDEL receptor inhibitor protein"  
 FT misc\_feature 154..279  
 FT /tag= d  
 FT /note= "Rat COMP pentamerisation domain"  
 FT misc\_feature 310..381  
 FT /tag= e  
 FT /note= "Camel IgG linker domain"  
 XX WO200006729-A1.  
 XX 10-FEB-2000.  
 XX 28-JUL-1999; 99WO-US17147.  
 XX 29-JUL-1998; 98US-0124671.  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 XX Roelman JE, Mayhew M, Hoe MH;  
 XX MPI: 2000-195296/17.  
 XX P-PSDB; AAY44967.  
 XX Inhibitors of the KDEL receptor which comprises an oligomerization  
 PT domain useful for promoting secretion of proteins which are normally  
 PT retained within the cell -  
 XX Disclosure; Fig 10; 87pp; English.  
 XX The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerisation domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a  
 CC cleavable signal peptide; a myc-tag; an N-glycosylation sequence; the  
 CC oligomerisation domain of rat cartilage oligomeric matrix protein  
 CC (COMP); a camel IgG linker domain and the carboxy-terminal sequence  
 CC KDEL. The subsequence G9CC is an alteration of rat COMP which provides  
 CC increased stability via disulphide bonds.  
 S0 Sequence 444 BP; 115 A; 122 C; 134 G; 73 T; 0 other;

Db 127 GGATCCAGCCTGGGTGAGACTGTTG-----CCACAGATGCTTCGAGAACTC 174  
 Qy 133 CAGAAACCAACGCGGCGCTGCAGAGACTGCGGACGCGAGCTCAAGGAG 192  
 Db 175 CAGAGACTAATGGGGCGCTGCAGACGTGAGAGAGCTTTCGACAGCAGTCAAGGAG 234  
 Qy 193 ATCAGTTCTCGAAAAACACGGTGATGAGAGTGAACGCGTGC----- 235  
 Db 235 ATCAGTTCTCGAAGAAATACGATATGATATGATACGCTTCGGAATGACCGCGACGC 294  
 Qy 236 -----GCGCGACGCGCGACCGGAAACCGGACCGCGACCGCGACCGCG 282  
 Db 295 ACCCGCGTACTACTCCGACGCGCGACCGGAAACCGGACCGCGACCGCGACCGCG 354  
 Qy 283 AACCCGACGCGGAAACCGGAAACCGGAAAGTTCCGATCATAGAAAAAGATGTTGAG 342  
 Db 355 AACCCGACGCGGAAACCGGAAACCGGAAAGTTCCGATCATAGAAAAAGATGTTGAG 414  
 Qy 343 GCGGCGGAGAAATTCATATGATCATCTCGAG 372  
 Db 415 GCGGCGGAGAAATTCATATGATCATCTCGAG 444  
 RESULT 9  
 AA250494  
 ID AA250494 standard; DNA; 357 BP.  
 AC AA250494;  
 DT 23-MAY-2000 (first entry)  
 XX KDEL receptor inhibitor-3 DNA.  
 XX KDEL receptor inhibitor; heat shock protein; immune response;  
 XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;  
 XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
 XX infectious disease; allergy; autoimmune disease; ss.  
 OS Chimeric - Mus sp.  
 OS Chimeric - Camelus sp.  
 XX Key Location/Qualifiers  
 FT CDS 10..327  
 FT /tag= a  
 FT /product= "KDEL receptor inhibitor protein"  
 FT sig\_peptide 10..68  
 FT /tag= b  
 FT /note= "Derived from mouse Bip"  
 FT mat\_peptide 69..324  
 FT /tag= c  
 FT /product= "Mature KDEL receptor inhibitor protein"  
 FT misc\_feature 97..222  
 FT /tag= d  
 FT /note= "Mouse TSP3 domain"  
 FT misc\_feature 223..294  
 FT /tag= e  
 FT /note= "Camel IgG linker domain"  
 XX WO200006729-A1.  
 XX 10-FEB-2000.  
 XX 28-JUL-1999; 99WO-US17147.  
 XX 29-JUL-1998; 98US-0124671.  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 XX Roelman JE, Mayhew M, Hoe MH;  
 XX MPI: 2000-195296/17.  
 XX P-PSDB; AAY44960.

XX Inhibitors of the KDEL receptor which comprises an oligomerisation  
 PT domain useful for promoting secretion of proteins which are normally  
 PR retained within the cell -  
 XX  
 PS Disclosure; Fig 3; 87pp; English.

XX The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerisation domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence encodes KDEL receptor inhibitor comprising regions encoding a  
 CC cleavable signal peptide; the oligomerisation domain of mouse  
 CC thrombospondin 3 (TSP3) trimerisation domain; a camel IgG linker domain  
 CC and the carboxy-terminal sequence KDEL. The subsequence GPCC is an  
 CC alteration of rat cartilage oligomeric matrix protein which provides  
 CC increased stability via disulphide bonds.  
 XX

SQ Sequence 357 BP; 84 A; 100 C; 114 G; 59 T; 0 other;

Query Match 47.7%; Score 177.6; DB 21; Length 357;  
 Best Local Similarity 73.4%; Pred. No. 3.8e-33;  
 Matches 2/3; Conservative 0; Mismatches 84; Indels 15; Gaps 3;

QY 1 AAGCTTACCATTGGAAGATGATGATTTAGGCTTGCCTTGGCGAGCTTCAGAC 60  
 DB 1 AAGCTTACCATTGGAAGATGATGATTTAGGCTTGCCTTGGCGAGCTTCAGAC 59  
 QY 61 GCTGCCAAAAGATGATGATGATTTAGGCTTGCCTTGGCGAGCTTCAGAC 120  
 DB 60 GCGGCGC--CGAGGATGATGATGATTTAGGCTTGCCTTGGCGAGCTTCAGAC 114  
 QY 121 CTTCGGGAAGTCCAGGAACCAAGCGCGGCTGCGAGAGCTGCGGAGCTGCGGCGAG 180  
 DB 115 -----CTACCCCTCTTCAACCAAGATGATGATTTAGGCTTGGCGAGCTTCAGAC 165  
 QY 181 CAGGTCAAGGAGATGATGATTTAGGCTTGCCTTGGCGAGCTTCAGAC 240  
 DB 166 CAGGTCAAGGAGATGATGATTTAGGCTTGCCTTGGCGAGCTTCAGAC 225  
 QY 241 CAGCGCGAGCGGAGATGATGATTTAGGCTTGCCTTGGCGAGCTTCAGAC 300  
 DB 226 CAGCGCGAGCGGAGATGATGATTTAGGCTTGCCTTGGCGAGCTTCAGAC 285  
 QY 301 GAACCGGAAGTACCGGATCATGAAAAAGATGATTTAGGCGGCGGAGATTCAT 360  
 DB 286 GAACCGGAAGTACCGGATCATGAAAAAGATGATTTAGGCGGCGGAGATTCAT 345  
 QY 361 ATGCATCTCGAG 372  
 DB 346 ATGCATCTCGAG 357

RESULT 10  
 ID AAA47734 standard; DNA; 755 BP.  
 XX AAA47734;  
 AC  
 XX 08-NOV-2000 (first entry)  
 DE Human COMP/TSP-1 chimeric protein coding sequence.  
 XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;  
 KW thrombospondin; angiogenesis; tumour; treatment; cancer;  
 XX

KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;  
 KW glaucoma; ds.  
 XX Homo sapiens.  
 OS Synthetic.  
 OS  
 FH Key Location/Qualifiers  
 FT CDS 26..754  
 FT /tag= a  
 FT /product= COMP/TSP-1 chimeric protein  
 PN WO200044908-A2.  
 XX  
 XX 03-AUG-2000.  
 XX  
 XX 01-FEB-2000; 2000WO-US02482.  
 XX  
 XX 01-FEB-1999; 99US-0118053.  
 XX  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX  
 XX Lawler JW;  
 XX  
 XX WPI: 2000-514823/46.  
 XX  
 XX P-PSDB; AAB00040.  
 DR  
 DR Nucleic acids encoding chimeric proteins such as cartilage oligomeric  
 PT matrix protein (COMP)/thrombospondin (TSP)-1 and 2, useful for  
 PT inhibiting angiogenesis and treating diseases such as cancer  
 XX  
 XX Claim 1; Fig 4a-b; 40pp; English.

XX New nucleic acids are described which encode a protein comprising  
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,  
 CC but not the TGF (transforming growth factor)-beta activation region  
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing  
 CC the second and third type-1 repeats and the COMP (cartilage  
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was  
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1  
 CC caused inhibition of the growth of tumours in mice models.  
 CC Thus the nucleic acids and proteins may be useful for treating  
 CC angiogenesis related diseases such as cancer (by reducing the rate of  
 CC growth and size of tumours), arthritis, psoriasis, diabetic  
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be  
 CC used for treating human immunodeficiency virus (HIV) infection.  
 CC Anti-angiogenic therapy has little toxicity, does not require the  
 CC therapeutic agent to enter tumour cells or cross the blood-brain  
 CC barrier, controls tumour growth independently of growth of  
 CC tumour cell heterogeneity, and does not induce drug resistance.  
 CC  
 XX  
 SQ Sequence 755 BP; 147 A; 245 C; 236 G; 127 T; 0 other;

Query Match 38.3%; Score 142.6; DB 21; Length 755;  
 Best Local Similarity 97.3%; Pred. No. 9.8e-25;  
 Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGCGGCGGAGATGCTTCCGGAAGTCAAGAAACCAACCGCGCTGCAG 157  
 DB 105 GCTCAGACCTGCGGCGGAGATGCTTCCGGAAGTCAAGAAACCAACCGCGCTGCAG 164  
 QY 158 ACGTGGGAGTCTGCTGCGGAGAGTCAAGAGATCAGTCTTGAACACGCTGA 217  
 DB 165 ACGTGGGAGTCTGCTGCGGAGAGTCAAGAGATCAGTCTTGAACACGCTGA 224  
 QY 218 TGGAGTGTGACGCGTGGCGGCGGAGCGG 246  
 DB 225 TGGAGTGTGACGCGTGGCGGAGTGCAGCG 253

RESULT 11  
 ID AAA47735 standard; DNA; 925 BP.  
 XX AAA47735;  
 XX



AC AAA47735;  
 XX 08-NOV-2000 (first entry)  
 XX  
 DE Human COMP/TSP-2 chimeric protein coding sequence.  
 XX  
 XX TSP-1; TSP-2; COMP; cartilage oligomeric matrix protein;  
 KW thrombospondin; angiogenesis; tumour; treatment; cancer;  
 KW arthritis; psoriasis; diabetic retinopathy; corneal graft rejection;  
 KW glaucoma; ds.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX  
 FT Key Location/Qualifiers  
 FT CDS 26..925  
 FT /\*tag= a  
 FT /product= COMP/TSP-2 chimeric protein  
 XX  
 XX MO200044908-A2.  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 01-FEB-2000; 2000WO-US02482.  
 XX  
 PR 01-FEB-1999; 99US-0118053.  
 XX  
 XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
 XX  
 XX Lawler JW;  
 XX  
 XX WPI; 2000-514823/46.  
 DR P-PSDB; AAB00041.  
 XX  
 PT Nucleic acid encoding chimeric proteins such as cartilage oligomeric  
 PT matrix protein (COMP)/thrombospondin (TSP)-1 and 2, useful for  
 PT inhibiting angiogenesis and treating diseases such as cancer  
 XX  
 PS Claim 44; Fig 5a-b; 40pp; English.  
 XX  
 CC New nucleic acid are described which encode a protein comprising  
 CC the second and third type 1 repeats of human TSP (thrombospondin)-1,  
 CC but not the TGF (transforming growth factor)-beta activation region,  
 CC of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing  
 CC the second and third type-1 repeats and the COMP (cartilage  
 CC oligomeric matrix protein) assembly sequence (COMP/TSP-1) was  
 CC produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1  
 CC caused inhibition of the growth of tumours in mice models.  
 CC Thus the nucleic acids and proteins may be useful for treating  
 CC angiogenesis related diseases such as cancer (by reducing the rate of  
 CC growth and size of tumours), arthritis, psoriasis, diabetic  
 CC retinopathy, corneal graft rejection, and glaucoma. They may also be  
 CC used for treating human immunodeficiency virus (HIV) infection.  
 CC Anti-angiogenic therapy has little toxicity, does not require the  
 CC therapeutic agent to enter tumour cells or cross the blood-brain  
 CC barrier, controls tumour growth independently of growth of  
 CC tumour cell heterogeneity, and does not induce drug resistance.  
 XX  
 XX Sequence 925 BP; 156 A; 310 C; 312 G; 147 T; 0 other;  
 SQ  
 Query Match 38.3%; Score 142.6; DB 21; Length 925;  
 Best Local Similarity 97.3%; Pred. No. 1e-24;  
 Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACTGCAGAAACCAACGCGCGCTGCAGG 157  
 DB 105 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACTGCAGAAACCAACGCGCGCTGCAGG 164  
 QY 158 ACGTGGCGGAGCTGGCTGGCGGAGGTCAGGAGATATCACTTCTGAAAAACACGGTGA 217  
 DB 165 ACGTGGCGGAGCTGGCTGGCGGAGGTCAGGAGATATCACTTCTGAAAAACACGGTGA 224  
 QY 218 TGAAGTGTGACGCGTGGCGGCGCGAGCCG 246

Db. 225 TGAAGTGTGACGCGTGGCGGAGATGCAGCAG 253  
 |||||  
 RESULT 12  
 ID ABR10887 standard; cDNA; 2439 BP.  
 XX ABR10887;  
 AC ABR10887;  
 XX  
 DT 04-DEC-2002 (first entry)  
 XX  
 XX Human breast cancer associated coding sequence SEQ ID NO: 1021.  
 XX  
 XX Human breast specific gene; breast cancer; differential expression;  
 KW cytoskeletal; gene therapy; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX MO200259271-A2.  
 XX  
 PD 01-AUG-2002.  
 XX  
 PF 25-JAN-2002; 2002WO-US02176.  
 XX  
 PR 25-JAN-2001; 2001US-263757P.  
 PR 25-APR-2001; 2001US-286090P.  
 PR 23-MAY-2001; 2001US-292517P.  
 XX  
 XX (GENE-) GENE LOGIC INC.  
 XX  
 XX Orr MS, Nation M, Diggans JC, Zeng W;  
 XX  
 XX WPI; 2002-674803/72.  
 DR  
 XX  
 PT Diagnosing breast cancer in a patient comprises detecting the level of  
 PT gene expression in cell or tissue samples, where a differential gene  
 PT expression is indicative of breast cancer -  
 XX  
 PS Claim 1; SEQ ID NO 1021; 260pp + Sequence Listing; English.  
 XX  
 CC The present invention relates to methods of diagnosing breast cancer in a  
 CC patient, which comprises detecting the level of expression in a tissue  
 CC sample of two or more genes selected from those shown in ABR09867-  
 CC ABR1112, where a differential expression of the genes indicates breast  
 CC cancer. The methods are useful in diagnosing, treating, detecting the  
 CC progression, and in monitoring treatment of breast cancer in patients.  
 CC The methods are also useful as a screening tool for agents that modulate  
 CC the onset or progression of breast cancer. The breast cancer genes may be  
 CC used as diagnostic markers for the prediction or identification of the  
 CC malignant state of breast tissue, for confirming the type and progression  
 CC of cancer, and for drug screening and assays. The present sequence is a  
 CC coding sequence of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub.published\_pct\_sequences.  
 XX  
 XX Sequence 2439 BP; 503 A; 758 C; 809 G; 369 T; 0 other;  
 SQ  
 Query Match 38.3%; Score 142.6; DB 24; Length 2439;  
 Best Local Similarity 97.3%; Pred. No. 1.2e-24;  
 Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACTGCAGAAACCAACGCGCGCTGCAGG 157  
 DB 105 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACTGCAGAAACCAACGCGCGCTGCAGG 164  
 QY 158 ACGTGGCGGAGCTGGCTGGCGGAGGTCAGGAGATATCACTTCTGAAAAACACGGTGA 217  
 DB 165 ACGTGGCGGAGCTGGCTGGCGGAGGTCAGGAGATATCACTTCTGAAAAACACGGTGA 224  
 QY 218 TGAAGTGTGACGCGTGGCGGCGCGAGCCG 246



XX (AVALON-) AVALON PHARM.  
PA Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI; 2002-188264/24.  
XX  
PT Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
XX  
XX Claim 1; SEQ ID 441; 44pp; English.  
XX  
CC The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilms' tumour.  
XX  
SO Sequence 2439 BP; 503 A; 758 C; 809 G; 369 T; 0 other;  
XX  
XX Query Match 38.3%; Score 142.6; DB 24; Length 2439;  
XX Best Local Similarity 97.3%; Pred. No. 1.2e-24;  
XX Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
XX  
OY 98 GTTCGACCTGGGCGCGGAGAGCTTCGGGAACTGCAGAAACCAACCGCGGCTGCAGG 157  
DB 105 GCTCAGACCTGGGCGCGGAGAGCTTCGGGAACTGCAGAAACCAACCGCGGCTGCAGG 164  
OY 158 ACGTCGGGAGCTGCTGCGGCGAGCAGTCTCAGGAGATCACTTCTGAAAAACAGCGTGA 217  
DB 165 ACGTCGGGAGCTGCTGCGGCGAGCAGTCTCAGGAGATCACTTCTGAAAAACAGCGTGA 224  
OY 218 TGGAGTGTGACCGCTGCGGCGCGGAGCCG 246  
DB 225 TGGAGTGTGACCGCTGCGGCGGAGTGCAGCAG 253  
XX  
XX RESULT 15  
XX ABL62870  
XX ID ABL62870 standard; DNA; 2439 BP.  
XX  
XX ABL62870;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Breast cancer related gene sequence SEQ ID NO:1207.  
XX  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;  
KW cytostatic; gene therapy; anti-neoplastic; Wilms' tumour; adenocarcinoma;  
KW gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX MO200194629-A2.  
XX  
XX 13-DEC-2001.  
XX  
XX 30-MAY-2001; 2001WO-US10838.  
XX  
PF

XX  
PR 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.  
PR 28-SEP-2000; 2000US-236111P.  
PR 29-SEP-2000; 2000US-236842P.  
PR 29-SEP-2000; 2000US-236891P.  
PR 02-OCT-2000; 2000US-237172P.  
PR 02-OCT-2000; 2000US-237173P.  
PR 02-OCT-2000; 2000US-237278P.  
PR 02-OCT-2000; 2000US-237294P.  
PR 02-OCT-2000; 2000US-237295P.  
PR 02-OCT-2000; 2000US-237316P.  
PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237599P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237606P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244867P.  
PR 01-NOV-2000; 2000US-245084P.  
XX  
XX (AVALON-) AVALON PHARM.  
XX  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
XX Soppet DR, Weaver Z;  
XX  
XX WPI; 2002-188264/24.  
XX  
XX  
XX Screening for anti-neoplastic agent involves exposing cells to a  
XX chemical agent to be tested for anti-neoplastic activity, and  
XX determining a change in expression of a gene of a signature gene set -  
XX  
XX Claim 1; SEQ ID 1207; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an  
XX anti-neoplastic agent. The method involves exposing cells to a chemical  
XX agent to be tested for anti-neoplastic activity, determining a change in  
XX expression of at least one gene (I) of a signature gene set, where (I)  
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
XX to ABL70110), or is at least 95% identical to (S), where a change in  
XX expression is indicative of anti-neoplastic activity. (I) has cytostatic  
XX activity and can be used in gene therapy. M1 can be used for screening  
XX an anti-neoplastic agent, and can be used for producing a product which  
XX is the data collected with respect to the anti-neoplastic agent as a  
XX result of M1, and the data is sufficient to convey the chemical  
XX structure and/or properties of the agent. M1 can be used in the  
XX treatment of cancer such as colon, breast, stomach, lung, thyroid,  
XX oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
XX adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
XX infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine

CC .carcinoma, papillary carcinoma and Wilm's tumour.

**XX**

Query Match	Score	DB	Length
38.3%	142.6;	24;	2439;
97.3%	Prod	No	1 39-24.

Best Local Similarity 97.3%; Pred. No. 1.2e-24; Matches 145; Conservative 0; Mismatches 4

Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

98 GTTCAGACCTGGGCCCGCAGATGCTTCGGAACTGCAGGAACCAACGCCGGCGTGCAGG 157

Db 105 GCTCAGACCTGGCCCGCAGATGCTTCGGAACTGCAGGAACCAACGCGGCTGCAGG 164

QY 158 ACGTGGGACTGGCTGCCGACAGGTCAGGAGATCACGTTCCGAAAAACACGGTGA 217

Db 165 ACGTCCGGA CTGGCTCCGCGCAGCAGGTCA GGGAGATCACGTTCTGAAAAACACCGTGA 224

218 TGGAGTGTGACGCGTGCGGGCCGACGCCG 246

Db 225 TGGAGTGTGACGCGGTGCGGATGCAGCAG 253

Search completed: August 26, 2003, 16:08:06  
Job time : 209 secs

Job time : 209 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: August 26, 2003, 15:10:25 ; Search time 55 Seconds  
(without alignments)  
2985.355 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 372

Sequence: 1 aagcttaccatgggaagta.....attccatgcatcgcag 372

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 113956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysts of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	372	100.0	372	US-09-124-671-24	Sequence 24, Appl
2	271.2	72.9	372	US-09-124-671-30	Sequence 30, Appl
3	264.8	71.2	372	US-09-124-671-28	Sequence 28, Appl
4	202	54.3	369	US-09-124-671-22	Sequence 22, Appl
5	194	52.2	369	US-09-124-671-20	Sequence 20, Appl
6	193.4	52.0	387	US-09-124-671-14	Sequence 14, Appl
7	192	51.6	387	US-09-124-671-16	Sequence 16, Appl
8	190.8	51.3	444	US-09-124-671-35	Sequence 35, Appl
9	177.6	47.7	357	US-09-124-671-18	Sequence 18, Appl
10	156.4	42.0	315	US-09-124-671-26	Sequence 26, Appl
11	56.4	15.2	480	PCT-US96-04648-1	Sequence 1, Appl
12	55.4	14.9	397	US-09-253-691-3	Sequence 3, Appl
13	55	14.8	1037	US-09-181-585-3	Sequence 3, Appl
14	55	14.8	1159	US-09-181-585-1	Sequence 1, Appl
15	55	14.8	1471	US-09-181-585-2	Sequence 1, Appl
16	53	14.2	521	US-09-643-597-334	Sequence 324, App
17	53	14.2	521	US-09-480-884A-324	Sequence 324, App
18	53	14.2	521	US-09-542-615A-324	Sequence 324, App
19	53	14.2	521	US-09-606-421B-324	Sequence 324, App
20	52.6	14.1	477	US-09-135-994-1	Sequence 1, Appl
21	52.6	14.1	477	US-09-684-843A-1	Sequence 1, Appl
22	52.2	14.0	23187	US-09-499-523-1	Sequence 1, Appl
23	51.4	13.8	234	US-08-469-802B-3	Sequence 3, Appl
24	51.4	13.8	234	US-08-267-803B-3	Sequence 3, Appl
25	50.4	13.5	71	US-08-464-318-11	Sequence 11, Appl
26	50.4	13.5	71	US-08-471-341-11	Sequence 11, Appl
27	50.4	13.5	71	US-08-461-566-11	Sequence 11, Appl

28	49.8	13.4	336	2	US-07-814-220-4	Sequence 4, Appl
29	49.8	13.4	336	2	US-07-812-421-4	Sequence 4, Appl
30	49.6	13.3	10348	2	US-08-457-273B-41	Sequence 41, Appl
31	49.6	13.3	10348	3	US-08-556-419-13	Sequence 13, Appl
32	49.6	13.3	10348	3	US-09-041-886-14	Sequence 14, Appl
33	49.6	13.3	10366	1	US-08-246-982A-5	Sequence 5, Appl
34	49.6	13.3	10366	1	US-08-453-265-5	Sequence 5, Appl
35	49.2	13.2	203	3	US-09-043-303-7	Sequence 7, Appl
36	48.4	13.0	71	1	US-08-464-318-12	Sequence 12, Appl
37	48.4	13.0	71	2	US-08-471-341-12	Sequence 12, Appl
38	48.4	13.0	71	2	US-08-461-566-12	Sequence 12, Appl
39	48.2	13.0	336	2	US-07-814-220-3	Sequence 3, Appl
40	48.2	13.0	336	2	US-07-812-421-3	Sequence 3, Appl
41	47.8	12.8	90	1	US-08-464-318-8	Sequence 8, Appl
42	47.8	12.8	90	2	US-08-471-341-8	Sequence 8, Appl
43	47.8	12.8	90	2	US-08-461-566-8	Sequence 8, Appl
44	47.4	12.7	2294	4	US-09-086-663A-70	Sequence 70, Appl
45	47.4	12.7	3334	4	US-09-086-663A-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-09-124-671-24  
Sequence 24, Application US/09124671A  
Patent No. 6160088  
GENERAL INFORMATION:  
APPLICANT: Rothman, James  
APPLICANT: Mayhew, Mark  
TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS  
FILE REFERENCE: 31488  
CURRENT APPLICATION NUMBER: US/09/124,671A  
CURRENT FILING DATE: 1998-07-29  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 24  
LENGTH: 372  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chimeric human COMP-KDEL  
US-09-124-671-24

DP?

Query Match	100.0%	Score 372, DB 3, Length 372;
Best Local Similarity	100.0%	Pred. No. 3e-81;
Matches 372; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	AAGCTTACCATGGGAGGATCATGATTTAGGCTTGGCCCTTGGGAGTGTGACG 60
DB	1	AAGCTTACCATGGGAGGATCATGATTTAGGCTTGGCCCTTGGGAGTGTGACG 60
QY	61	GCTGCCAAAAAAGATCCAGCTGGGTGAGAGCTTTGTCAGACTGTGGGCCAGATG 120
DB	61	GCTGCCAAAAAAGATCCAGCTGGGTGAGAGCTTTGTCAGACTGTGGGCCAGATG 120
QY	121	CTTGGGAACTGCAGAAACCAACGCGGCGCTGCAGAGAGCTGCGGAGCTGCGCAG 180
DB	121	CTTGGGAACTGCAGAAACCAACGCGGCGCTGCAGAGAGCTGCGGAGCTGCGCAG 180
QY	181	CAGTCCAGGAGATCAGCTTCTTAAAAAACAAGTGTGAGTGTGACGCTGTGGGGCGG 240
DB	181	CAGTCCAGGAGATCAGCTTCTTAAAAAACAAGTGTGAGTGTGACGCTGTGGGGCGG 240
QY	241	CAGCCGACGCGAAACCGCAGCCGCGGAGCGCGGAGCGCGAAACCGCAGCCGAAACCG 300
DB	241	CAGCCGACGCGAAACCGCAGCCGCGGAGCGCGGAGCGCGGAGCGCGAAACCGCAGCCGAAACCG 300
QY	301	GAACCGAAGGTACCGATCATCAAGAAAGATGATTGTTAGCGGCGCGCAAAATTCAT 360
DB	301	GAACCGAAGGTACCGATCATCAAGAAAGATGATTGTTAGCGGCGCGCAAAATTCAT 360

QY 361 ATGCATCTCGAG 372  
| | | | | | | | | |  
| | | | | | | | | |  
Db 361 ATGCATCTCGAG 372

## RESULT 2

US-09-124-671-30

/ Sequence 30, Application US/09124671A

/ Patent No. 6160088

/ GENERAL INFORMATION:

/ APPLICANT: Rothman, James

/ APPLICANT: Mayhew, Mark

/ APPLICANT: Hoe, Mee

/ TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS

/ FILE REFERENCE: 31488

/ CURRENT APPLICATION NUMBER: US/09/124,671A

/ CURRENT FILING DATE: 1998-07-29

/ NUMBER OF SEQ ID NOS: 42

/ SOFTWARE: FastSeq for Windows Version 3.0

/ SEQ ID NO 30

/ LENGTH: 372

/ TYPE: DNA

/ ORGANISM: Artificial Sequence

/ FEATURE:

/ OTHER INFORMATION: chimeric human TSP4-KDEL

US-09-124-671-30

Query Match 72.9%; Score 271.2; DB 3; Length 372;

Best Local Similarity 83.1%; Pred. No. 6,1e-57;

Matches 309; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 AAGCTTACCATGGAGGATCATGATTTAGGCTTGTGCTGCTTGTGCGGAGTCTGCAGC 60  
| | | | | | | | | |  
| | | | | | | | | |  
Db 1 AAGCTTACCATGGAGGATCATGATTTAGGCTTGTGCTGCTTGTGCGGAGTCTGCAGC 60  
| | | | | | | | | |  
| | | | | | | | | |  
QY 61 GGTGCAAAAAGATCCAGCTGGTGGAGACTGTTCAGACCTGGCCCGCAGATG 120  
| | | | | | | | | |  
| | | | | | | | | |  
Db 61 GGTGCAAAAAGATCCAGCTGGTGGAGACTGTTCAGACCTGGCCCGCAGATG 120  
| | | | | | | | | |  
| | | | | | | | | |  
QY 121 CTTGCGGAAGTCAAGAAACCAACGCGCGCTGCAGAGCTGCGGAGTGTGCGGCGG 180  
| | | | | | | | | |  
| | | | | | | | | |  
Db 121 TTTGGCTCAAAATACACAAATTAACCACTCTGCGAGAGGTAAAGACCTTTGTAGACAG 180  
| | | | | | | | | |  
| | | | | | | | | |  
QY 181 CAGGTCAGGAGATCATGCTTCTGAAAAACACGCGTGATGAGTGTGACGCGTGCAGCGG 240  
| | | | | | | | | |  
| | | | | | | | | |  
Db 181 CAGGTTAAGAAACATCATTTTGGCAAAACCATATGCTGAATGCGAGGCTTGCAGTCCG 240  
| | | | | | | | | |  
| | | | | | | | | |  
QY 241 CAGCGCGACGCGAAACCGGACGCGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 300  
| | | | | | | | | |  
| | | | | | | | | |  
Db 241 CAGCGCGACGCGAAACCGGACGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
| | | | | | | | | |  
| | | | | | | | | |  
QY 301 GAACCGGAAGTACCGGATCATCAGAAAAAGATGATTGTAGCGCGCGCAGAAATTCAT 360  
| | | | | | | | | |  
| | | | | | | | | |  
Db 301 GAACCGGAAGTACCGGATCATCAGAAAAAGATGATTGTAGCGCGCGCAGAAATTCAT 360  
| | | | | | | | | |  
| | | | | | | | | |  
QY 361 ATGCATCTCGAG 372  
| | | | | | | | | |  
| | | | | | | | | |  
Db 361 ATGCATCTCGAG 372  
| | | | | | | | | |  
| | | | | | | | | |

## RESULT 3

US-09-124-671-28

/ Sequence 28, Application US/09124671A

/ Patent No. 6160088

/ GENERAL INFORMATION:

/ APPLICANT: Rothman, James

/ APPLICANT: Mayhew, Mark

/ APPLICANT: Hoe, Mee

/ TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS

/ FILE REFERENCE: 31488

/ CURRENT APPLICATION NUMBER: US/09/124,671A

/ CURRENT FILING DATE: 1998-07-29

/ NUMBER OF SEQ ID NOS: 42

/ SOFTWARE: FastSeq for Windows Version 3.0

/ SEQ ID NO 28

/ LENGTH: 372

/ TYPE: DNA

/ ORGANISM: Artificial Sequence

/ FEATURE:

/ OTHER INFORMATION: chimeric human TSP3-KDEL

US-09-124-671-28

Query Match 71.2%; Score 264.8; DB 3; Length 372;

Best Local Similarity 82.0%; Pred. No. 2,1e-55;

Matches 305; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 AAGCTTACCATGGAGGATCATGATTTAGGCTTGTGCTGCTTGTGCGGAGTCTGCAGC 60  
| | | | | | | | | |  
| | | | | | | | | |  
Db 1 AAGCTTACCATGGAGGATCATGATTTAGGCTTGTGCTGCTTGTGCGGAGTCTGCAGC 60  
| | | | | | | | | |  
| | | | | | | | | |  
QY 61 GGTGCAAAAAGATCCAGCTGGTGGAGACTGTTCAGACCTGGCCCGCAGATG 120  
| | | | | | | | | |  
| | | | | | | | | |  
Db 61 GGTGCAAAAAGATCCAGCTGGTGGAGACTGTTCAGACCTGGCCCGCAGATG 120  
| | | | | | | | | |  
| | | | | | | | | |  
QY 121 CTTGCGGAAGTCAAGAAACCAACGCGCGCTGCAGAGCTGCGGAGTGTGCGGCGG 180  
| | | | | | | | | |  
| | | | | | | | | |  
Db 121 CTCACCCAGCTCACCCTTTCAACCAAGATCCTAGTGAAGCTTGGAGCAGATCCGAGAC 180  
| | | | | | | | | |  
| | | | | | | | | |  
QY 181 CAGGTCAGGAGATCATGCTTCTGAAAAACACGCGTGATGAGTGTGACGCGTGCAGCGG 240  
| | | | | | | | | |  
| | | | | | | | | |  
Db 181 CAGGTCAGGAGATCATGCTTCTGAAAAACACGCGTGATGAGTGTGACGCGTGCAGCGG 240  
| | | | | | | | | |  
| | | | | | | | | |  
QY 241 CAGCGCGACGCGAAACCGGACGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGG 300  
| | | | | | | | | |  
| | | | | | | | | |  
Db 241 CAGCGCGACGCGAAACCGGACGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
| | | | | | | | | |  
| | | | | | | | | |  
QY 301 GAACCGGAAGTACCGGATCATCAGAAAAAGATGATTGTAGCGCGCGCAGAAATTCAT 360  
| | | | | | | | | |  
| | | | | | | | | |  
Db 301 GAACCGGAAGTACCGGATCATCAGAAAAAGATGATTGTAGCGCGCGCAGAAATTCAT 360  
| | | | | | | | | |  
| | | | | | | | | |  
QY 361 ATGCATCTCGAG 372  
| | | | | | | | | |  
| | | | | | | | | |  
Db 361 ATGCATCTCGAG 372  
| | | | | | | | | |  
| | | | | | | | | |

## RESULT 4

US-09-124-671-22

/ Sequence 22, Application US/09124671A

/ Patent No. 6160088

/ GENERAL INFORMATION:

/ APPLICANT: Rothman, James

/ APPLICANT: Mayhew, Mark

/ APPLICANT: Hoe, Mee

/ TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS

/ FILE REFERENCE: 31488

/ CURRENT APPLICATION NUMBER: US/09/124,671A

/ CURRENT FILING DATE: 1998-07-29

/ NUMBER OF SEQ ID NOS: 42

/ SOFTWARE: FastSeq for Windows Version 3.0

/ SEQ ID NO 22

/ LENGTH: 369

/ TYPE: DNA

/ ORGANISM: Artificial Sequence

/ FEATURE:

/ OTHER INFORMATION: chimeric Xenopus laevis TSP4-KDEL

US-09-124-671-22

Query Match 54.3%; Score 202; DB 3; Length 369;

Best Local Similarity 79.1%; Pred. No. 3e-40; Indels 0; Gaps 0;

Matches 239; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 71 AAGATCCAGGCTGGTGAAGACTGTTTCAACTGCGCCCGCAGATGCTTGGGAGAC 130  
| | | | | | | | | |  
| | | | | | | | | |  
Db 68 AGGATCCAGGCTGGTGAAGACTGTTTGTGTGACGTCAGACAGATGATTGGCCAGA 127  
| | | | | | | | | |  
| | | | | | | | | |  
QY 131 TGACAGAAACCAACGCGGCGCTGCAGAGACTGCGGAGCTGCGGCGAGCAGAGTCAAGG 190  
| | | | | | | | | |  
| | | | | | | | | |

RESULT 6  
US-09-124-671-14

Query Match	51.6%	Score 192;	DB 3;	Length 387;
Best Local Similarity	76.4%;	Pred. NO. 7.8e-38;		

Matches 307; Conservative 0; Mismatches 50; Indels 45; Gaps 4;

QY 1 AAGCTTACCATGGAGAGTACATGATTTTAAAGCTTGTCTGCGCTTGTGCGGAGCTGTGCAGC 60  
Db 1 AAGCTTACCATGGAGAGTACATGATTTTAAAGCTTGTCTGCGCTTGTGCGGAGCTGTGCAGC 59  
QY 61 GCTGCCAAAAGATCCAGCTGGGTGAGAGCTGTTGTTCAAGACTGTGGCCCGCAGATG 120  
Db 60 GCGGGC--CGAGGGATCCAGCTGGGTGAGAGCTGTTGT-----CCACAGATG 105  
QY 121 CTTCCGGAAGCTGAGGAAACCAACGCGGCTGCAAGAGCTGCGGAGCTGGCTGCGGAG 180  
Db 106 CTTCCGGAAGCTGAGGAAACCAACGCGGCTGCAAGAGCTGCGGAGCTGGCTGCGGAG 165  
QY 181 CAGGTCAGGAGATCACTGTTCTGAAAAAACAAGGTGAGTGTGACCGCTGCG----- 235  
Db 166 CAGGTCAGGAGATCACTGTTCTGAAAAAACAAGGTGAGTGTGACCGCTGCGGAAATG 225  
QY 236 -----GCGCGACGCGGACCGGAAACCGGACCGGACCGG 270  
Db 226 CAGCGCGACGACCGGACCGGAAACCGGAAACCGGAAACCGGACCGGACCGGACCGG 285  
QY 271 CAGCGCGACGCGGAAACCGGACCGGAAACCGGAAACCGGAAACCGGAAACCGGAAAC 330  
Db 286 CAGCGCGACGCGGAAACCGGACCGGAAACCGGAAACCGGAAACCGGAAACCGGAAAC 345  
QY 331 GATGAGTTGTAGCGCGCGCGCAGAAATTCATATGATCTCGAG 372  
Db 346 GATGAGTTGTAGCGCGCGCGCAGAAATTCATATGATCTCGAG 387

RESULT 8  
US-09-124-671-35  
; Sequence 35, Application US/09124671A  
; Patent No. 6160088  
; GENERAL INFORMATION:  
; APPLICANT: Rothman, James  
; APPLICANT: Mayhew, Mark  
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS  
; FILE REFERENCE: 31488  
; CURRENT APPLICATION NUMBER: US/09/124,671A  
; CURRENT FILING DATE: 1998-07-29  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 35  
; LENGTH: 444  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: KDEL-myc  
US-09-124-671-35

Query Match 51.3%; Score 190.8; DB 3; Length 444;  
Best Local Similarity 80.6%; Pred. No. 1.6e-37;  
Matches 266; Conservative 0; Mismatches 22; Indels 42; Gaps 2;

QY 73 GATCCAGCCTGGGTGAGAGATGTTGTCAGACCTGGCCCGCAGATGCTTGGGAACTG 132  
Db 127 GGATCCAGCCTGGGTGAGAGATGTTGT-----CCACAGATGCTTGGGAACTG 174  
QY 133 CAGGAAACCAACGCGGCTGCAAGAGCTGCGGAGCTGGCTGCGGACGAGTCAAGGAG 192  
Db 175 CAGGAGACTAAATGCGGCGCTGCAAGAGCTGAGAGAGCTTTCGACAGAGGTCAGAGAG 234  
QY 193 ATCAGTTCCTGAAAAACAACGCTGATGAGTGTGACGCTGG----- 235  
Db 235 ATCAGTTCCTGAAAAACAACGCTGATGAGTGTGACGCTGGAGATGAGCCCGCAGCC 294  
QY 236 -----GCGCGACGCGGACCGGAAACCGGACCGGACCGGACCGGACCGG 282  
Db 295 ACCCGCGGTAAGTCCGACGCGCAGCGGAAACCGGACCGGACCGGACCGGACCGG 354

QY 283 AAACCGGACCGGAAACCGGAAACCGGAAACCGGAAACCGGAAACCGGAAACCGGAAAC 342  
Db 355 AAACCGGACCGGAAACCGGAAACCGGAAACCGGAAACCGGAAACCGGAAACCGGAAAC 414  
QY 343 GCGGCGGAGAAATTCATATGATCTCGAG 372  
Db 415 GCGGCGGAGAAATTCATATGATCTCGAG 444

RESULT 9  
US-09-124-671-18  
; Sequence 18, Application US/09124671A  
; Patent No. 6160088  
; GENERAL INFORMATION:  
; APPLICANT: Rothman, James  
; APPLICANT: Mayhew, Mark  
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS  
; FILE REFERENCE: 31488  
; CURRENT APPLICATION NUMBER: US/09/124,671A  
; CURRENT FILING DATE: 1998-07-29  
; NUMBER OF SEQ ID NOS: 42  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 18  
; LENGTH: 357  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chimeric mouse TSP1-KDEL  
US-09-124-671-18

Query Match 47.7%; Score 177.6; DB 3; Length 357;  
Best Local Similarity 73.4%; Pred. No. 2.3e-34;  
Matches 273; Conservative 0; Mismatches 84; Indels 15; Gaps 3;

QY 1 AAGCTTACCATGGAGAGTACATGATTTTAAAGCTTGTCTGCGCTTGTGCGGAGCTGTGCAGC 60  
Db 1 AAGCTTACCATGGAGAGTACATGATTTTAAAGCTTGTCTGCGCTTGTGCGGAGCTGTGCAGC 59  
QY 61 GCTGCCAAAAGATCCAGCTGGGTGAGAGCTGTTGTTCAAGACTGTGGCCCGCAGATG 120  
Db 60 GCGGGC--CGAGGATCCAGCTGGGTGAGAGCTGTTGTTCAAGACTGTGGCCCGCAGATG 114  
QY 121 CTTCCGGAAGCTGAGGAAACCAACGCGGCTGCAAGAGCTGCGGAGCTGGCTGCGGAG 180  
Db 115 -----CTCACCTCTTCAACCAAGATCTTAAGTGAAGCTTCCGAGACGATCCGAGAC 165  
QY 181 CAGGTCAGGAGATCACTGTTCTGAAAAACAACGCTGATGAGTGTGACGCTGGGCGG 240  
Db 166 CAGGTCAGGAGATCACTGTTCTGAAAAACAACGCTGATGAGTGTGACGCTGGGCGG 225  
QY 241 CAGCGCGACGCGGAAACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGAAACCG 300  
Db 226 CAGCGCGACGCGGAAACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGAAACCG 285  
QY 301 GAACCGGAGGTATCCGATCATCAGAAAAAGATGATGTTGAGCGCGCGCAGAAATTCAT 360  
Db 286 GAACCGGAGGTATCCGATCATCAGAAAAAGATGATGTTGAGCGCGCGCAGAAATTCAT 345  
QY 361 ATGATCTCGAG 372  
Db 346 ATGATCTCGAG 357

RESULT 10  
US-09-124-671-26  
; Sequence 26, Application US/09124671A  
; Patent No. 6160088  
; GENERAL INFORMATION:  
; APPLICANT: Rothman, James  
; APPLICANT: Mayhew, Mark  
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS



RESULT 11  
PCT-US96-04648-1  
Sequence 1, Application PC/TUS9604648  
GENERAL INFORMATION:  
APPLICANT: Cell Genesys, Inc.  
TITLE OF INVENTION: Transplantation of Genetically Modified Cells Having Low Level  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 941114187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PCDOS/MSDOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/04648  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sherwood, Pamela J  
REGISTRATION NUMBER: 36,677  
REFERENCE/DOCKET NUMBER: Cell 23-1

```

RESULT 13
US-09-181-585-3/c
; Sequence 3, Application US/09181585
; Patent No. 6524791
; GENERAL INFORMATION:
; APPLICANT: Rannu, Laura P.W.
; APPLICANT: Koob, Michael
; TITLE OF INVENTION: SPINOCREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION
; FILE REFERENCE: 11000900101
; CURRENT APPLICATION NUMBER: US/09/181,585

```

/ CURRENT FILING DATE: 1998-10-28  
/ NUMBER OF SEQ ID NOS: 18  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO: 3  
/ LENGTH: 1037  
/ TYPE: DNA  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: Description of Artificial Sequence: cDNA  
/ OTHER INFORMATION: comprising exons E, C, and A  
US-09-181-585-3

Query Match 14.8%; Score 55; DB 4; Length 1037;  
Best Local Similarity 53.5%; Pred. No. 0.0001;  
Matches 115; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 94 TGTGTTGACAGCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACCGCGCGCTG 153  
DB 938 TATTTTAAAAAATGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 869  
QY 154 CAGGACGTGCGGAGCTGCTGCGGACAGACAGACAGACAGACAGACAGACAGACAG 213  
DB 868 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 809  
QY 214 GTGATGAGTGTGACGCGTGCAGGCGCCGACGCGCGGAAACCGCAGCGCGCAG 273  
DB 808 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 749  
QY 274 CCGCAGCCGAAACCGCAGCCGAAACCGGAAACCGGA 308  
DB 748 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 714

RESULT 14  
US-09-181-585-1/C  
/ Sequence 1, Application US/09181585  
/ Patent No. 6524791  
/ GENERAL INFORMATION:  
/ APPLICANT: Rannum, Laura P. W.  
/ APPLICANT: Koob, Michael  
/ TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION  
/ FILE REFERENCE: 11000900101  
/ CURRENT APPLICATION NUMBER: US/09/181,585  
/ CURRENT FILING DATE: 1998-10-28  
/ NUMBER OF SEQ ID NOS: 18  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO: 1  
/ LENGTH: 1159  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-09-181-585-1

Query Match 14.8%; Score 55; DB 4; Length 1159;  
Best Local Similarity 53.5%; Pred. No. 0.0001;  
Matches 115; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 94 TGTGTTGACAGCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACCGCGCGCTG 153  
DB 740 TATTTTAAAAAATGACAGACAGACAGACAGACAGACAGACAGACAGACAGCAG 681  
QY 154 CAGGACGTGCGGAGCTGCTGCGGACAGACAGACAGACAGACAGACAGACAGACAG 213  
DB 680 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 621  
QY 214 GTGATGAGTGTGACGCGTGCAGGCGCCGACGCGCGGAAACCGCAGCGCGCAG 273  
DB 620 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 561  
QY 274 CCGCAGCCGAAACCGCAGCCGAAACCGGAAACCGGA 308  
DB 560 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 526

RESULT 15  
US-09-181-585-2/C  
/ Sequence 2, Application US/09181585  
/ Patent No. 6524791  
/ GENERAL INFORMATION:  
/ APPLICANT: Rannum, Laura P. W.  
/ APPLICANT: Koob, Michael  
/ TITLE OF INVENTION: SPINOCEREBELLAR ATAXIA TYPE 8 AND METHODS OF DETECTION  
/ FILE REFERENCE: 11000900101  
/ CURRENT APPLICATION NUMBER: US/09/181,585  
/ CURRENT FILING DATE: 1998-10-28  
/ NUMBER OF SEQ ID NOS: 18  
/ SOFTWARE: PatentIn Ver. 2.0  
/ SEQ ID NO: 2  
/ LENGTH: 1471  
/ TYPE: DNA  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: Description of Artificial Sequence: cDNA  
/ OTHER INFORMATION: comprising exons D, C, B, and A  
US-09-181-585-2

Query Match 14.8%; Score 55; DB 4; Length 1471;  
Best Local Similarity 53.5%; Pred. No. 0.00011;  
Matches 115; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 94 TGTGTTGACAGCTGGGCCCCGAGATGCTTCGGGAACTGCAGGAAACCAACCGCGCGCTG 153  
DB 1362 TATTTTAAAAAATGACAGACAGACAGACAGACAGACAGACAGACAGACAGCAG 1303  
QY 154 CAGGACGTGCGGAGCTGCTGCGGACAGACAGACAGACAGACAGACAGACAGACAG 213  
DB 1302 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1243  
QY 214 GTGATGAGTGTGACGCGTGCAGGCGCCGACGCGCGGAAACCGCAGCGCGCAG 273  
DB 1242 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1183  
QY 274 CCGCAGCCGAAACCGCAGCCGAAACCGGAAACCGGA 308  
DB 1182 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1148

Search completed: August 26, 2003, 15:35:31  
Job time : 56 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2003, 15:34:34 ; Search time 805 Seconds

(without alignments)  
1038.903 Million cell updates/sec

Title: US-09-696-872-24

Sequence: 1 aagcttaccatgggaagta.....aattccatcgcctcgcag 372

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA:  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09C\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142.6	38.3	755	US-09-919-603-4	Sequence 4, Appl1
2	142.6	38.3	925	US-09-919-603-6	Sequence 6, Appl1
3	142.6	38.3	2439	US-09-954-531-140	Sequence 140, App
4	142.6	38.3	2439	US-09-954-531-359	Sequence 359, App
5	142.6	38.3	2439	US-09-918-6248-49	Sequence 49, Appl
6	142.6	38.3	2439	US-10-301-822-40	Sequence 40, Appl
7	142.6	38.3	2439	US-10-177-293-71	Sequence 71, Appl
8	142.6	38.3	2439	US-09-111-911-5	Sequence 5, Appl1
9	142.6	38.3	2439	US-09-782-378A-3	Sequence 3, Appl1
10	142.6	38.3	2439	US-10-156-761-4994	Sequence 4994, Ap
11	142.6	38.3	2439	US-09-918-995-4559	Sequence 4559, Ap
12	142.6	38.3	2439	US-09-918-995-4560	Sequence 4560, Ap
13	142.6	38.3	2439	US-09-918-995-4560	Sequence 4560, Ap
14	142.6	38.3	2439	US-09-918-995-4560	Sequence 4560, Ap
15	142.6	38.3	2439	US-09-918-995-4560	Sequence 4560, Ap
16	142.6	38.3	2439	US-09-918-995-4560	Sequence 4560, Ap

17	53	14.2	521	12	US-10-117-982-324	Sequence 324, App
18	53	14.2	1082	9	US-09-925-301-441	Sequence 441, App
19	53	14.2	1412	2	US-10-301-822-154	Sequence 154, App
20	53	14.2	3747	13	US-10-071-766-33	Sequence 33, Appl
21	52.6	14.1	41936	10	US-09-867-768A-116	Sequence 116, App
22	51.2	13.8	88421	10	US-10-080-797-14	Sequence 14, Appl
23	51.2	13.8	88421	10	US-09-976-059-1	Sequence 1, Appl1
24	50.6	13.6	9025608	14	US-10-156-761-1	Sequence 1, Appl1
25	49.6	13.3	614	14	US-10-215-432-30	Sequence 30, Appl
26	49.6	13.3	614	14	US-10-215-432-31	Sequence 31, Appl
27	48.2	13.0	1080	14	US-10-215-432-26	Sequence 26, Appl
28	47.8	12.8	372	13	US-10-079-623-187	Sequence 187, App
29	46.4	12.5	1926	12	US-10-294-804-3	Sequence 3, Appl1
30	46.4	12.5	2614	11	US-09-822-846-491	Sequence 491, App
31	46.4	12.5	6604	10	US-09-880-107-1748	Sequence 1748, Ap
32	46.4	12.5	8705	14	US-10-291-230-14	Sequence 14, Appl
33	46.4	12.5	8705	14	US-10-291-249-14	Sequence 14, Appl
34	46.2	12.4	2276	12	US-10-137-870-9	Sequence 9, Appl1
35	46.2	12.4	2276	12	US-10-140-018-9	Sequence 9, Appl1
36	46.2	12.4	2276	12	US-10-140-021-9	Sequence 9, Appl1
37	46.2	12.4	2276	12	US-10-140-274-9	Sequence 9, Appl1
38	46.2	12.4	2276	12	US-10-140-471-9	Sequence 9, Appl1
39	46.2	12.4	2276	12	US-10-140-807-9	Sequence 9, Appl1
40	46.2	12.4	2276	12	US-10-140-922-9	Sequence 9, Appl1
41	46.2	12.4	2276	12	US-10-140-924-9	Sequence 9, Appl1
42	46.2	12.4	2276	12	US-10-140-926-9	Sequence 9, Appl1
43	46.2	12.4	2276	12	US-10-141-698-9	Sequence 9, Appl1
44	46.2	12.4	2276	12	US-10-141-702-9	Sequence 9, Appl1
45	46.2	12.4	2276	12	US-10-141-704-9	Sequence 9, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-919-603-4  
Sequence 4, Application US/09919603  
Patent No. US20020137679A1  
GENERAL INFORMATION:  
APPLICANT: Lawler, John W.  
TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP  
FILE REFERENCE: 1440.1033-007  
CURRENT APPLICATION NUMBER: US/09/919, 603  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: PCT/US00/02482  
PRIOR FILING DATE: 2000-02-01  
PRIOR APPLICATION NUMBER: 60/118,053  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 755  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: fusion gene  
US-09-919-603-4  
Query Match 38.3%; Score 142.6; DB 10; Length 755;  
Best Local Similarity 97.3%; Pred. No. 3.7e-32;  
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 98 GTTCAGACCTGGGCGCCGAGATGCTTGGGAACTGCAGAGAAACCAACGCGCGCTGCAGG 157  
DB 105 GCTCAGACCTGGGCGCCGAGATGCTTGGGAACTGCAGAGAAACCAACGCGCGCTGCAGG 164  
QY 158 ACGGCGGAGTGGCTGGCGAGAGAGTCAAGGAGATCACTTCTGAAACACGGTGA 217  
DB 165 ACGGCGGAGTGGCTGGCGAGAGAGTCAAGGAGATCACTTCTGAAACACGGTGA 224  
QY 218 TGGAGTGTGACGCGTGGCGGCGCGAGCGG 246

```

Db      225  TGGAGTGTGACGCGCTGCCGGATGCACGACG  253

RESULT 2
US-09-919-603-6
: Sequence 6, Application US/09919603
: Patent No. US20020137679A1
: GENERAL INFORMATION:
:   APPLICANT: Lawler, John W.
:   TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
:   TITLE OF INVENTION: Chimeric Proteins
:   FILE REFERENCE: 1440.1033-007
:   CURRENT APPLICATION NUMBER: US/09/919,603
:   CURRENT FILING DATE: 2001-07-30
:   PRIOR APPLICATION NUMBER: PCT/US00/02482
:   PRIOR FILING DATE: 2000-02-01
:   PRIOR APPLICATION NUMBER: 60/118,053
:   PRIOR FILING DATE: 1999-02-01
:   NUMBER OF SEQ ID NOS: 21
:   SOFTWARE: FastSeq for Windows Version 4.0
:   SEQ ID NO 6
:   LENGTH: 925
:   TYPE: DNA
:   ORGANISM: Artificial Sequence
:   FEATURE:
:   OTHER INFORMATION: fusion gene
: US-09-919-603-6

Query Match      38.3%; Score 142.6; DB 10; Length 925;
Best Local Similarity 97.3%; Pred. No. 3.8e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      98  GTTCAGACTGTGGGCCCGACAGATGCTTCGGGAAGTGCAGAAACCAACGCGCGCTGCAGG  157
Db      105  GCTCAGACTGTGGGCCCGACAGATGCTTCGGGAAGTGCAGAAACCAACGCGCGCTGCAGG  164

QY      158  ACGTGGCGGACTGCTGCTGCGCGACGACAGTCAAGGAGATCAGTCTCTGAAAAACGCGTGA  217
Db      165  ACGTGGCGGACTGCTGCTGCGCGACGACAGTCAAGGAGATCAGTCTCTGAAAAACGCGTGA  224

QY      218  TGGAGTGTGACGCGCTGCCGGCGCGACGCG  246
Db      225  TGGAGTGTGACGCGCTGCCGGCGATGCACGACG  253

RESULT 3
US-09-954-531-140
: Sequence 140, Application US/09954531
: Patent No. US20020165180A1
: GENERAL INFORMATION:
:   APPLICANT: Weaver, Zoe
:   TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cancer
:   TITLE OF INVENTION: Gene Sets
:   FILE REFERENCE: 689290-77
:   CURRENT APPLICATION NUMBER: US/09/954,531
:   CURRENT FILING DATE: 2002-05-02
:   PRIOR APPLICATION NUMBER: US/60/233,133
:   PRIOR FILING DATE: 2000-09-18
:   PRIOR APPLICATION NUMBER: US/60/234,009
:   PRIOR FILING DATE: 2000-09-20
:   PRIOR APPLICATION NUMBER: US/60/234,034
:   PRIOR FILING DATE: 2000-09-20
:   PRIOR APPLICATION NUMBER: US/60/234,509
:   PRIOR FILING DATE: 2000-09-22
:   PRIOR APPLICATION NUMBER: US/60/234,567
:   PRIOR FILING DATE: 2000-09-22
:   NUMBER OF SEQ ID NOS: 1392
:   SOFTWARE: PatentIn version 3.0
:   SEQ ID NO 140
:   LENGTH: 2439
:   TYPE: DNA
:   ORGANISM: Homo sapiens
: US-09-954-531-140

```

```

Query Match      38.3%; Score 142.6; DB 10; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4,6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      98 GTTCAGACCTGGGCCCGCAGATGCTTCGGGAACCTGCAGAAACCAACGCGCGCTGCAGG 157
      |||
DB      105 GCTCAGACCTGGGCCCGCAGATGCTTCGGGAACCTGCAGAAACCAACGCGCGCTGCAGG 164

QY      158 ACGTCGGGACTGCGCTGCGGCAGCAGATCAGGAGATCACTTCTGAAAAACACGGTGA 217
      |||
DB      165 ACGTCGGGACTGCGCTGCGGCAGCAGATCAGGAGATCACTTCTGAAAAACACGGTGA 224

QY      218 TGGAGTGTGACCGCTGTCGGGGCCGCAGCCG 246
      |||
DB      225 TGGAGTGTGACCGCTGTCGGGATGCAGCAG 253

RESULT 4
US-09-954-531-359
; Sequence 359, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cancer
; TITLE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 359
; LENGTH: 2439
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-359

Query Match      38.3%; Score 142.6; DB 10; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4,6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      98 GTTCAGACCTGGGCCCGCAGATGCTTCGGGAACCTGCAGAAACCAACGCGCGCTGCAGG 157
      |||
DB      105 GCTCAGACCTGGGCCCGCAGATGCTTCGGGAACCTGCAGAAACCAACGCGCGCTGCAGG 164

QY      158 ACGTCGGGACTGCGCTGCGGCAGCAGATCAGGAGATCACTTCTGAAAAACACGGTGA 217
      |||
DB      165 ACGTCGGGACTGCGCTGCGGCAGCAGATCAGGAGATCACTTCTGAAAAACACGGTGA 224

QY      218 TGGAGTGTGACCGCTGTCGGGGCCGCAGCCG 246
      |||
DB      225 TGGAGTGTGACCGCTGTCGGGATGCAGCAG 253

RESULT 5
US-09-918-624B-49
; Sequence 49, Application US/09918624B
; Publication No. US20030113720A1
; GENERAL INFORMATION:
; APPLICANT: Schedye, Xiao Min
; APPLICANT: Sornasee, Thierly
; TITLE OF INVENTION: CDNAS EXPRESSED IN ADIPOCYTE DIFFERENTIATION
; FILE REFERENCE: PA-0033 US

```

```

CURRENT APPLICATION NUMBER: US/09/918,624B
CURRENT FILING DATE: 2002-12-03
PRIOR APPLICATION NUMBER: 60/222,470
PRIOR FILING DATE: 2000-07-28
NUMBER OF SEQ ID NOS: 71
SOFTWARE: PERL Program
SEQ ID NO 49
LENGTH: 2439
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: GenBank ID No. US20030113720A1 9602449
US-09-918-624B-49

Query Match      38.3%; Score 142.6; DB 11; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4.6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAAGAAACCAACGCGGCTGCAGG 157
    |||||
DB 105 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAAGAAACCAACGCGGCTGCAGG 164
    |||||

QY 158 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCACGTTCTGAAAAACAACGGTGA 217
    |||||
DB 165 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCACGTTCTGAAAAACAACGGTGA 224
    |||||

QY 218 TGGAGTGTGACGCGTGCAGGCGCGGCGCGG 246
    |||||
DB 225 TGGAGTGTGACGCGTGCAGGCGCGGCGCGG 253
    |||||

RESULT 6
US-10-301-822-40
Sequence 40, Application US/10301822
Publication No. US20030148410A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Berger, Allison
APPLICANT: Guillemette, Tracy L.
APPLICANT: Kametkar, Shubhangi
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Thibodeau, Stephen N.
APPLICANT: Burgart, Lawrence J.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
FILE REFERENCE: MEMO1-029P2RMM
CURRENT APPLICATION NUMBER: US/10/301,822
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US 60/339,971
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 60/361,978
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/381,988
PRIOR FILING DATE: 2002-05-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 40
LENGTH: 2439
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (26) ... (2299)
US-10-301-822-40

Query Match      38.3%; Score 142.6; DB 12; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4.6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAAGAAACCAACGCGGCTGCAGG 157
    |||||
DB 105 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAAGAAACCAACGCGGCTGCAGG 164
    |||||

QY 158 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCACGTTCTGAAAAACAACGGTGA 217
    |||||
DB 165 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCACGTTCTGAAAAACAACGGTGA 224
    |||||

QY 218 TGGAGTGTGACGCGTGCAGGCGCGGCGCGG 246
    |||||
DB 225 TGGAGTGTGACGCGTGCAGGCGCGGCGCGG 253
    |||||
```

```

CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 71
LENGTH: 2439
TYPE: DNA
ORGANISM: Homo sapiens
US-10-177-293-71

Query Match      38.3%; Score 142.6; DB 14; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4.6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAAGAAACCAACGCGGCTGCAGG 157
    |||||
DB 105 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAAGAAACCAACGCGGCTGCAGG 164
    |||||

QY 158 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCACGTTCTGAAAAACAACGGTGA 217
    |||||
DB 165 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCACGTTCTGAAAAACAACGGTGA 224
    |||||

QY 218 TGGAGTGTGACGCGTGCAGGCGCGGCGCGG 246
    |||||
DB 225 TGGAGTGTGACGCGTGCAGGCGCGGCGCGG 253
    |||||

RESULT 7
US-10-177-293-71
Sequence 71, Application US/10177293
Publication No. US20030124128A1
GENERAL INFORMATION:
APPLICANT: Lilly, James
APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Ganavarrpu, Manjula
APPLICANT: Kametkar, Shubhangi
APPLICANT: Mertens, Maureen
APPLICANT: Meyer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bast Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Puzstai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 71
LENGTH: 2439
TYPE: DNA
ORGANISM: Homo sapiens
US-10-177-293-71

Query Match      38.3%; Score 142.6; DB 14; Length 2439;
Best Local Similarity 97.3%; Pred. No. 4.6e-32;
Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAAGAAACCAACGCGGCTGCAGG 157
    |||||
DB 105 GCTCAGACCTGGGCGCCGAGATGCTTCGGGAACCTGCAAGAAACCAACGCGGCTGCAGG 164
    |||||

QY 158 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCACGTTCTGAAAAACAACGGTGA 217
    |||||
DB 165 ACGTCCGGGACTGGCTGCGGAGCAGGTCAAGAGATCACGTTCTGAAAAACAACGGTGA 224
    |||||

QY 218 TGGAGTGTGACGCGTGCAGGCGCGGCGCGG 246
    |||||
DB 225 TGGAGTGTGACGCGTGCAGGCGCGGCGCGG 253
    |||||
```

Db 225 TGGAGTGTGACGCGTGGGAGATGACAGCAG 253

RESULT 8  
US-09-111-911-5/c

/ Sequence 5, Application US/09111911  
/ Publication No. US20030096768A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Mold, William S.M.  
/ TITLE OF INVENTION: Inhibiting Apoptosis With Adenovirus R1D Protein  
/ FILE REFERENCE: 16153-5587  
/ CURRENT FILING DATE: 1998-07-08  
/ NUMBER OF SEQ ID NOS: 5  
/ SOFTWARE: Patent In Ver. 2.0  
/ SEQ ID NO 5  
/ LENGTH: 34427  
/ TYPE: DNA  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: Description of Artificial Sequence: Plasmid  
/ OTHER INFORMATION: Combining E.coli and Adenovirus Sequences  
US-09-111-911-5

Query Match 15.6%; Score 58; DB 11; Length 34427;  
Best Local Similarity 92.4%; Pred. No. 5e-07;  
Matches 61; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 CCATGGGAAGGTACATGATTTTAGGCTTGCTGCGCCCTTGCGGAGTCTGCAGCGCTGCCA 67  
Db 2656 CCAGATGATGATGATGATTTAGGCTTGCTGCGCCCTTGCGGAGTCTGCAGCGCTGCCA 2597

QY 68 AAAAG 73  
Db 2596 AAAAG 2591

RESULT 9  
US-09-782-378A-3

/ Sequence 3, Application US/09782378A  
/ Patent No. US20020102731A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Hearing, Patrick  
/ APPLICANT: Bahou, Wadie  
/ APPLICANT: Sandalon, Ziv  
/ APPLICANT: Gnatenko, Dmitrii  
/ TITLE OF INVENTION: Adenoviral Vectors  
/ FILE REFERENCE: STONYB-04970  
/ CURRENT APPLICATION NUMBER: US/09/782,378A  
/ CURRENT FILING DATE: 2001-02-12  
/ PRIOR APPLICATION NUMBER: 60/237,747  
/ PRIOR FILING DATE: 2000-10-02  
/ NUMBER OF SEQ ID NOS: 27  
/ SOFTWARE: Patent In version 3.0  
/ SEQ ID NO 3  
/ LENGTH: 35937  
/ TYPE: DNA  
/ ORGANISM: Human adeno-associated virus 2  
US-09-782-378A-3

Query Match 15.6%; Score 58; DB 10; Length 35937;  
Best Local Similarity 92.4%; Pred. No. 5.1e-07;  
Matches 61; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 8 CCATGGGAAGGTACATGATTTTAGGCTTGCTGCGCCCTTGCGGAGTCTGCAGCGCTGCCA 67  
Db 28807 CCAGATGATGATGATGATTTAGGCTTGCTGCGCCCTTGCGGAGTCTGCAGCGCTGCCA 28866

QY 68 AAAAG 73  
Db 28867 AAAAG 28872

RESULT 10  
US-10-156-761-4994

/ Sequence 4994, Application US/10156761  
/ Publication No. US20030119018A1  
/ GENERAL INFORMATION:  
/ APPLICANT: OMURA, SATOSHI  
/ APPLICANT: IKEDA, HARUO  
/ APPLICANT: ISHIKAWA, JUN  
/ APPLICANT: HORIKAWA, HIROSHI  
/ APPLICANT: SHIBA, TADAYOSHI  
/ APPLICANT: SAKAKI, YOSHIYUKI  
/ APPLICANT: HATTORI, MASAHIRA  
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
/ FILE REFERENCE: 249-262  
/ CURRENT APPLICATION NUMBER: US/10/156,761  
/ CURRENT FILING DATE: 2002-05-29  
/ PRIOR APPLICATION NUMBER: JP 2001-204089  
/ PRIOR FILING DATE: 2001-05-30  
/ PRIOR APPLICATION NUMBER: JP 2001-272697  
/ PRIOR FILING DATE: 2001-08-02  
/ NUMBER OF SEQ ID NOS: 15109  
/ SEQ ID NO 4994  
/ LENGTH: 1734  
/ TYPE: DNA  
/ ORGANISM: Streptomyces avermitilis  
/ FEATURE:  
/ NAME/KEY: CDS  
/ LOCATION: (1)..(1734)  
US-10-156-761-4994

Query Match 15.5%; Score 57.8; DB 14; Length 1734;  
Best Local Similarity 89.9%; Pred. No. 3.1e-07;  
Matches 62; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 237 GCCGAGCCGCGAAGCCGGAACCGCGAGCCGCGAGCCGCGAGCCGCGAAGCCGCGAA 296  
Db 105 GCCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGCA 164

QY 297 ACCGGAACC 305  
Db 165 GCCGCAACC 173

RESULT 11  
US-10-156-761-1

/ Sequence 1, Application US/10156761  
/ Publication No. US20030119018A1  
/ GENERAL INFORMATION:  
/ APPLICANT: OMURA, SATOSHI  
/ APPLICANT: IKEDA, HARUO  
/ APPLICANT: ISHIKAWA, JUN  
/ APPLICANT: HORIKAWA, HIROSHI  
/ APPLICANT: SHIBA, TADAYOSHI  
/ APPLICANT: SAKAKI, YOSHIYUKI  
/ APPLICANT: HATTORI, MASAHIRA  
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
/ FILE REFERENCE: 249-262  
/ CURRENT APPLICATION NUMBER: US/10/156,761  
/ CURRENT FILING DATE: 2002-05-29  
/ PRIOR APPLICATION NUMBER: JP 2001-204089  
/ PRIOR FILING DATE: 2001-05-30  
/ PRIOR APPLICATION NUMBER: JP 2001-272697  
/ PRIOR FILING DATE: 2001-08-02  
/ NUMBER OF SEQ ID NOS: 15109  
/ SEQ ID NO 1  
/ LENGTH: 9025608  
/ TYPE: DNA  
/ ORGANISM: Streptomyces avermitilis  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (4187715)  
/ OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

Query Match 15.5%; Score 57.8; DB 14; Length 9025608;

Best Local Similarity 89.9%; Pred. No. 1.8e-06;

Matches 62; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 237 GCCGAGCCGCGAGCCGGAACCGCAGCCGCGAGCCGCGAGCCGGAACCGCAGCCGGA 296

Db 6065691 GCCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGAGCCGCGCA 6065750

QY 297 ACCGGAAC 305

Db 6065751 GCCGCAAC 6065759

RESULT 12

US-09-918-995-4559

Sequence 4559, Application US/09918995

Publication No. US20030073623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918,995

CURRENT FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US/09/235,076

PRIOR FILING DATE: 1999-01-20

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4559

LENGTH: 428

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)...(428)

OTHER INFORMATION: n = A,T,C or G

US-09-918-995-4559

Query Match 14.5%; Score 53.8; DB 11; Length 428;

Best Local Similarity 55.7%; Pred. No. 3.5e-06;

Matches 103; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 116 AGATGCTTCGGGAATGAGAAACCAACCGCGCGCTGACAGAGCTGGGAGTGGCTGC 175

Db 68 AGATGCTGAGGCGCATCTGCGGTCAATCCAGCGCGAGAGCAAGACCTGCTGC 127

QY 176 GCGAGAGTCAAGAGATCACTTCTGAAACACAGGTATGATGATGAGCGTGC 235

Db 128 AGCAGAGCCCCCTCGAGCGCGAGCGCGCAAGCGCTGAAAGAGGCGCTGAGAGAG 187

QY 236 GCGCGAGCGCGAGCGCGAAACCGCAGCGCGAGCGCGAGCGCGAGCGCGAAACCGCAGCGCA 295

Db 188 GCGAGCAACCCCGAGCGCGCAATCAACCCCGAGCGCGCAACCCCGAGCGCGCGCC 247

QY 296 AACCG 300

Db 248 AGCAG 252

RESULT 13

US-09-918-995-4560

Sequence 4560, Application US/09918995

Publication No. US20030073623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918,995

CURRENT FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US/09/235,076

PRIOR FILING DATE: 1999-01-20

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4560

LENGTH: 433

TYPE: DNA

ORGANISM: Homo sapiens

US-09-918-995-4560

Query Match 14.5%; Score 53.8; DB 11; Length 433;

Best Local Similarity 55.7%; Pred. No. 3.5e-06;

Matches 103; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 116 AGATGCTTCGGGAATGAGAAACCAACCGCGCGCTGACAGAGCTGGGAGTGGCTGC 175

Db 68 AGATGCTGAGGCGCATCTGCGGTCAATCCAGCGCGAGAGCAAGACCTGCTGC 127

QY 176 GCGAGAGTCAAGAGATCACTTCTGAAACACAGGTATGATGATGAGCGTGC 235

Db 128 AGCAGAGCCCCCTCGAGCGCGAGCGCGCAAGCGCTGAAAGAGGCGTGTGAGAGAG 187

QY 236 GCGCGAGCGCGAGCGCGAAACCGCAGCGCGAGCGCGAGCGCGAGCGCGAAACCGCAGCGCA 295

Db 188 GCGAGCAACCCCGAGCGCGCAATCAACCCCGAGCGCGCAACCCCGAGCGCGCGCC 247

QY 296 AACCG 300

Db 248 AGCAG 252

RESULT 14

US-09-735-705-324

Sequence 324, Application US/09735705

Patent No. US20020052329A1

GENERAL INFORMATION:

APPLICANT: Wang, Tongtong

APPLICANT: Fan, Liqun

APPLICANT: Kalos, Michael D.

APPLICANT: Bangur, Chaitanya S.

APPLICANT: Hosken, Nancy

APPLICANT: Fanger, Gary R.

APPLICANT: Li, Samuel X.

APPLICANT: Wang, Aijun

APPLICANT: Skelky, Yasir A.W.

APPLICANT: Henderson, Robert A.

APPLICANT: McNeill, Patricia D.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

FILE REFERENCE: 210121.455C14

CURRENT APPLICATION NUMBER: US/09/735,705

CURRENT FILING DATE: 2000-12-12

NUMBER OF SEQ ID NOS: 419

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 324

LENGTH: 521

TYPE: DNA

ORGANISM: Homo sapien

US-09-735-705-324

Query Match 14.2%; Score 53; DB 9; Length 521;

Best Local Similarity 58.6%; Pred. No. 6.3e-06;

Matches 92; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 149 CGCTGAGAGAGTGGGAGTGGCTGCGGAGCAGGTCAAGGAGATCACTTCTGAAA 208

Db 52 CGCTGAGAGTGTGACAGTAACAAATCGTCAAGGCGCATCTGCGGTCAATTCACGCGGC 111

QY 209 ACAGGTGATGAGTGTGAGCGCTGCGGCGCGAGCGCGAGCGCGAAACCGCAGCGCAGC 268

Db 112 AGAAGCAGCAGCAGCTGTGACAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 171

QY 269 GCGAGCGCAGCGCAACCGCAGCGCGAAACCGGAAAC 305





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 26, 2003, 15:10:25 ; Search time 1432 Seconds  
(without alignments)  
613.733 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 372  
Sequence: 1 aagcttaccatggaagta.....aatccatgcatctcgag 372

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_eebba: \*  
2: em\_eebba: \*  
3: em\_eebba: \*  
4: em\_eebba: \*  
5: em\_eebba: \*  
6: em\_eebba: \*  
7: em\_eebba: \*  
8: em\_eebba: \*  
9: em\_eebba: \*  
10: em\_eebba: \*  
11: em\_eebba: \*  
12: em\_eebba: \*  
13: em\_eebba: \*  
14: em\_eebba: \*  
15: em\_eebba: \*  
16: em\_eebba: \*  
17: em\_eebba: \*  
18: em\_eebba: \*  
19: em\_eebba: \*  
20: em\_eebba: \*  
21: em\_eebba: \*  
22: em\_eebba: \*  
23: em\_eebba: \*  
24: em\_eebba: \*  
25: em\_eebba: \*  
26: em\_eebba: \*  
27: em\_eebba: \*  
28: em\_eebba: \*  
29: em\_eebba: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match length	ID	Description
1	137.8	37.0	9	AT751993 cn13a10.x
2	137.8	37.0	12	BG900764 HOA42-1-H
3	137.8	37.0	12	BG897326 HOA12-1-E
4	137.8	37.0	12	BG896418 HOA30-1-F

5	137.8	37.0	370	12	BG897214	BG897214 HOA33-1-F
6	137.8	37.0	371	12	BG898406	BG898406 HOA8-1-F1
7	137.8	37.0	372	12	BG899727	BG899727 HOA32-1-A
8	137.8	37.0	373	12	BG896387	BG896387 HOA30-1-E
9	137.8	37.0	374	12	BG896380	BG896380 HOA30-1-E
10	137.8	37.0	375	9	AM068338	AM068338 cn20g10.Y
11	137.8	37.0	376	12	BG900556	BG900556 HOA42-1-C
12	137.8	37.0	377	12	BG898027	BG898027 HOA20-1-H
13	137.8	37.0	378	12	BG896332	BG896332 HOA37-1-F
14	137.8	37.0	379	12	BG900343	BG900343 HOA37-1-F
15	137.8	37.0	380	12	BG898382	BG898382 HOA8-1-F4
16	137.8	37.0	381	12	BG896503	BG896503 HOA30-1-C
17	137.8	37.0	382	12	BG899366	BG899366 HOA40-1-E
18	137.8	37.0	383	12	BG900455	BG900455 HOA41-1-C
19	137.8	37.0	384	12	BG897369	BG897369 HOA12-1-B
20	137.8	37.0	385	12	BG897511	BG897511 HOA14-1-A
21	137.8	37.0	386	12	BG898968	BG898968 HOA21-1-C
22	137.8	37.0	387	12	BG900465	BG900465 HOA41-1-C
23	137.8	37.0	388	12	BG898309	BG898309 HOA39-1-A
24	137.8	37.0	389	12	BG898604	BG898604 HOA19-1-G
25	137.8	37.0	390	12	BG899695	BG899695 HOA39-1-D
26	137.8	37.0	391	12	BG896715	BG896715 HOA49-1-D
27	137.8	37.0	392	12	BG897500	BG897500 HOA14-1-G
28	137.8	37.0	393	12	BG898326	BG898326 HOA55-1-C
29	137.8	37.0	394	12	BG898329	BG898329 HOA55-1-D
30	137.8	37.0	395	9	AT751616	AT751616 cn11a01.Y
31	137.8	37.0	396	12	BG898451	BG898451 HOA9-1-D1
32	137.8	37.0	397	9	AM067829	AM067829 cn19f09.X
33	137.8	37.0	398	12	BG898353	BG898353 HOA55-1-D
34	137.8	37.0	399	12	BG900134	BG900134 HOA51-1-D
35	137.8	37.0	400	12	BG898861	BG898861 HOA40-1-C
36	137.8	37.0	401	12	BG898320	BG898320 HOA55-1-C
37	137.8	37.0	402	12	BG897459	BG897459 HOA13-1-H
38	137.8	37.0	403	12	BG898924	BG898924 HOA40-1-F
39	137.8	37.0	404	12	BG896555	BG896555 HOA47-1-D
40	137.8	37.0	405	12	BG901017	BG901017 HOA52-1-G
41	137.8	37.0	406	12	BG901170	BG901170 HOA54-1-H
42	137.8	37.0	407	9	AU298537	AU298537 AU298537
43	137.8	37.0	408	12	BG899058	BG899058 HOA21-1-E
44	137.8	37.0	409	12	BG899411	BG899411 HOA45-1-A
45	137.8	37.0	410	12	BG899174	BG899174 HOA22-1-D

#### ALIGNMENTS

RESULT 1  
LOCUS AT751993 328 bp mRNA linear EST 20-JUN-2002  
DEFINITION cn13a10.x1 Normal Human Trabecular Bone Cells Homo sapiens CDNA  
clone NHTBC-cn13a10 random, mRNA sequence.

ACCESSION AT751993  
VERSION AT751993.1 GI:5130257

KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
Jia, L.B., Young, M.F., Touchman, J.W., Bouffard, G.G.,  
Beckstrom-Sternberg, S.M., Green, E.D., Powell, J.I., Yang, L.M., Robey  
P.G., Hotchkiss, R.N. and Francomano, C.A.  
SCAP: The Skeletal Genome Anatomy Project  
Unpublished  
Contact: Libin Jia  
Medical Genetics Branch  
National Human Genome Research Institute  
10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA  
Tel: 301-496-4877  
Fax: 301-496-7157  
Email: libin@helix.nih.gov  
DNA Sequencing and analyses by National Institutes of Health  
Intramural Sequencing Center (NISC).



RESULT 4	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT	FEATURES	ORIGIN
158	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
208	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
218	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
268	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
158	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
208	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
218	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
268	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
158	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
208	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
218	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
268	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
158	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
208	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
218	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
268	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
158	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
208	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
218	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
268	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
158	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
208	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
218	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
268	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-2001	BC896418													
158	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
208	ACGTCGGGAGCTGCTGTGGCGAGCAGCGGTCAAGAGATACGCTCTGAAAAACCGGTGA	370 bp mRNA linear EST 06-NOV-2001	BC896418													
218	TGAGAGTGTGACCGCGTGCAGGCGCGCAGCCG	370 bp mRNA linear EST 06-NOV-														

DEFINITION	HMOA3-1-F1 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA sequence.					
ACCESSION	BG897214 GI:14307455					
VERSION	EST.					
KEYWORDS	Homo sapiens (human)					
ORGANISM	Homo sapiens					
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 370) Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Sathie,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.					
TITLE	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage CDNA libraries Osteoarthr. Cartil. 9 (7), 641-653 (2001)					
JOURNAL MEDLINE PUBMED	21482651 11597177					
COMMENT	Contact: Sanjay Kumar UM2109 Glaxosmithkline 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA Tel: 610-270-7245 Fax: 610-270-5598 Email: sanjay_kumar-1@sk.com Seq primer: T7. Location/Qualifiers 1..370 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /tissue_type="cartilage" /lab_host="E.coli DH10 B" /clone_idb="HOA (Human Osteoarthritic Cartilage)" /note=Vector: pSPORT I; Site_1: SalI; Site_2: NotI; Directional"					
FEATURES	source					
BASE COUNT	67 a 127 c 120 g 56 t					
ORIGIN						
Query Match	37.0%; Score 137.8; DB 12; Length 370;					
Best Local Similarity	95.3%; Pred. No. 4,2e-22;					
Matches 142; Conservative	0; Mismatches 7; Indels 0; Gaps 0;					
Db	98 GTTTCAGACTGTGGGCCCGCAGATGCTTGCGGAATGTCAAGAAACCACGCGCTGCAGG 157 189 GCTAGAAGCTGGGGCCGCGCAGATGCTTGCGGAATGTCAAGAAACCAACGCGCGCTGCAGG 248					
Dc	158 ACGTCGGGAGACTGGCTGCGCGAGCAGCATGTCAAGGAGATCAAGTTCTGAAAAAACCGGTGA 217 249 ACGGCGCGGAGACTGCTCGCGCAGCAGGTCAGGAGATCACGTTCTGAAAAAACCGGTGA 308					
Dg	218 TGAGTGTGAACGCGTGGCGCGCGAGCG 246 309 TGAAGTGTGAACGCGTGGCGCGAGTGCAGAG 337					
RESULT 6	Bg898406 371 bp mRNA linear EST 06-NOV-2001					
LOCUS	HMOA-1-F10 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA sequence.					
ACCESSION	BG898406					
VERSION	BG898406.1 GI:14308655					
KEYWORDS	EST.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 371) Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Sathie,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Lark,M.W.					

TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
IDENTIFICATION AND INITIAL CHARACTERIZATION OF 5000 EXPRESSED SEQUENCED TAGS (ESTs) EACH FROM ADULT HUMAN NORMAL AND OSTEOARTHRITIC CARTILAGE CDNA LIBRARIES	Osteoarthritis. Cartil. 9 (7), 641-653 (2001)	21482651		
CONTACT: SANJAY KUMAR		11597177		
UM2109	GlaxoSmithKline			
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA				
Tel: 610-270-7245				
Fax: 610-270-5598				
Email: sanjay.kumar-1@gsk.com				
Seq primer: T7.				
Location/Qualifiers				
1..371				
/organism="Homo sapiens"				
/mol_type="mRNA"				
/db_xref="taxon:9606"				
/issue_type="cartilage"				
/lab_host="E. coli DH10 B"				
/clone_lib="HOA (Human Osteoarthritic Cartilage)"				
/note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI; Directional"				
BASE COUNT	65 a	126 c	118 g	62 t
ORIGIN				
Query Match	37.0%	Score 137.8;	DB 12;	Length 371;
Best Local Similarity	95.3%	Pred. No. 4.2e-22;		
Matches 142;	Conservative	0;	Mismatches 7;	Indels 0;
Gaps 0;				
QY	98	GTTTCAGACCTTGCGGCCGAGATGCTTCGGGAACTGCAGAAACCAACGCGCGCTGCAGG	157	
Db	157	GCTAGACACTTGCGGCCGAGATGCTTCGGGAACTGCAGAAACCAACGCGCGCTGCAGG	216	
QY	158	ACGTCGGCGGACTGCTGCGCGAGACGATCAGGAGATCACTTCCTGAAAAACCGGTGA	217	
Db	217	ACGTCGGCGGACTGCTGCGCGAGACGATCAGGAGATCACTTCCTGAAAAACCGGTGA	276	
QY	218	TGAGTGTGAACGCGTGCAGCGCGCGAGCCG	246	
Db	277	TGAGTGTGAACGCGTGCAGCGCGAGCGAG	305	
RESULT 7				
BG899727		372 bp	mRNA	linear
LOCUS	HOA32-1-A8 HOA	(Human Osteoarthritic Cartilage)	Homo sapiens cDNA,	
DEFINITION	mRNA sequence.			
ACCESSION	BG899727			
VERSION	BG899727.1	GI:14309992		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1. (bases 1 to 372)			
	Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Sathie,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Laik,M.W.			
	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries			
	Osteoarthr. Cartil. 9 (7), 641-653 (2001)			
TITLE	JOURNAL			
MEDLINE	21482651			
PUBMED	11597177			
COMMENT	Contact: Sanjay Kumar			
	UM2109			
	GlaxoSmithKline			
	709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA			
	Tel: 610-270-7245			
	Fax: 610-270-5598			
	Email:sanjay_kumar-1@gsk.com			

FEATURES		Seq primer: T7.	
source	location/Qualifiers		
	1..372		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/rname="cartilage"		
	/lab_host="E.coli DH10 B"		
	/clone_lib="HOA (Human Osteoarthritic Cartilage)"		
	/note="Vector: pSPORT I, Site_1: SalI, Site_2: NotI; Directional"		
BASE COUNT	68 a 130 c 117 g 57 t		
ORIGIN			
Query Match	37.0%; Score 137.8; DB: 12; Length 372;		
Best Local Similarity	95.3%; Prod. No. 4.2e-22;		
Matches 142;	Conservative 0; Mismatches 7; Indels 0; Gaps 0;		
QY	98 GTTCAGACCTGGGCGCCGACAGATGCTTGGGAACTGCAGAAACCAACGCGCGCTGCAGG 157		
Db	189 GCTCAGACTGTGGGCGCCGACATGCTTGGGAACTGCAGAAACCAACGCGCGCTGCAGG 248		
QY	158 ACGTGGCGGACTGGCTTGGCGGACAGGTCAAGGAGATCACTTCTTAAAAACCGGTGA 217		
Db	249 ACGGCGCGGACTGCTGGCGGACAGGTCAAGGAGATCACTTCTTAAAAACCGGTGA 308		
QY	218 TGGAGTGTGACGCGTGGCGGCGCCGACGCG 246		
Db	309 TGGAGTGTGACGCGTGGCGGCGCCGATGACGAG 337		
RESULT 8	BG896387 373 bp mRNA linear EST 06-NOV-2001		
LOCUS	BG896387		
DEFINITION	HOA30-1-87 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA, mRNA sequence.		
ACCESSION	BG896387		
VERSION	BG896387.1 GI:14306628		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryotic: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 373) Kumar,S., Connor,U.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J., Satche,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and Laik,M.W.		
TITLE	Identification and initial characterization of 5000 expressed sequenced tags (ESTs) each from adult human normal and osteoarthritic cartilage cDNA libraries		
JOURNAL	Osteoarthr. Cartil. 9 (7), 641-653 (2001)		
MEDLINE	21482651		
PUBMED	11597177		
COMMENT	Contact: Sanjay Kumar UW2109		
	GlaxoSmithKline		
	709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA		
	Tel: 610-270-7245		
	Fax: 610-270-5598		
	Email: sanjay.kumar-1@gsk.com		
	Seq primer: T7.		
FEATURES	location/Qualifiers		
source	1..373		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/rname="cartilage"		
	/lab_host="E.coli DH10 B"		
	/clone_lib="HOA (Human Osteoarthritic Cartilage)"		
	/note="Vector: pSPORT I, Site_1: SalI, Site_2: NotI; Directional"		
BASE COUNT	68 a 130 c 118 g 57 t		
ORIGIN			

Query Match 37.0%; Score 137.8; DB 12; Length 373;  
 Best Local Similarity 95.3%; Pred. No. 4.2e-22;  
 Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGCCCGCAGATGCTTCGGGAACCTGCAGAAACCAACCGCGCTGCAGG 157  
 DB 189 GCTCAGACCTGGCCCGCAGATGCTTCGGGAACCTGCAGAAACCAACCGCGCTGCAGG 248  
 QY 158 ACGTGGCGGACTGGCTGGCGAGCAGCTCAGGAGATCACTTCCTGAAAAACCGTGA 217  
 DB 249 ACGTGGCGGACTGGCTGGCGAGCAGCTCAGGAGATCACTTCCTGAAAAACCGTGA 308  
 QY 218 TGGAGTGTGACGCGTGCAGCGCGCAGCCG 246  
 DB 309 TGGAGTGTGACGCGTGCAGCGCGCAGCGAG 337

RESULT 9 BG896390 373 bp mRNA linear EST 06-NOV-2001  
 LOCUS HOA30-1-B10 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,  
 DEFINITION mRNA sequence.  
 ACCESSION BG896390  
 VERSION BG896390.1 GI:14306631  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 373)  
 Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,  
 Satche,G., Mu,J.P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and  
 Lack,M.W.  
 Identification and initial characterization of 5000 expressed  
 sequenced tags (ESTs) each from adult human normal and  
 osteoarthritic cartilage cDNA libraries  
 Osteoarthritis Cartil. 9 (7), 641-653 (2001)  
 21482651  
 11597177  
 COMMENT Contact: Sanjay Kumar  
 UW2109  
 GlaxoSmithKline  
 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA  
 Tel: 610-270-7245  
 Fax: 610-270-5598  
 Email: sanjay\_kumar-1@gsk.com  
 Seq primer: 17.  
 Location/Qualifiers  
 1..373  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /tissue\_type="cartilage"  
 /lab\_host="E.coli DH10 B"  
 /clone\_lib="HOA (Human Osteoarthritic Cartilage)"  
 /note="Vector: pSPORT 1; Site\_1: SalI; Site\_2: NotI;  
 Directional"  
 BASE COUNT 69 a 129 c 118 g 57 t  
 ORIGIN

Query Match 37.0%; Score 137.8; DB 12; Length 373;  
 Best Local Similarity 95.3%; Pred. No. 4.2e-22;  
 Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGCCCGCAGATGCTTCGGGAACCTGCAGAAACCAACCGCGCTGCAGG 157  
 DB 190 GCTCAGACCTGGCCCGCAGATGCTTCGGGAACCTGCAGAAACCAACCGCGCTGCAGG 249  
 QY 158 ACGTGGCGGACTGGCTGGCGAGCAGCTCAGGAGATCACTTCCTGAAAAACCGTGA 217  
 DB 250 ACGTGGCGGACTGGCTGGCGAGCAGCTCAGGAGATCACTTCCTGAAAAACCGTGA 309

QY 218 TGGAGTGTGACGCGTGCAGCGCCGAGCCG 246  
 DB 310 TGGAGTGTGACGCGTGCAGGATGCAGCAG 338

RESULT 10 AM068338 374 bp mRNA linear EST 20-JUN-2002  
 LOCUS c20g10.y1 Normal Human Trabecular Bone Cells Homo sapiens cDNA  
 DEFINITION clone NHTRC\_c20g10 random, mRNA sequence.  
 ACCESSION AM068338  
 VERSION AM068338.1 GI:6023336  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 374)  
 Jia,L.B., Young,M.F., Touchman,J.W., Bouffard,G.G.,  
 Beckstrom-Sternberg,S.M., Green,E.D., Powell,J.I., Yang,L.M., Robey  
 P.G., Hotchkiss,R.N. and Francomano,C.A.  
 SGAP: The Skeletal Genome Anatomy Project  
 Unpublished  
 TITLE JOURNAL  
 COMMENT Contact: Libin Jia  
 Medical Genetics Branch  
 National Human Genome Research Institute  
 10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA  
 Tel: 301-402-4877  
 Fax: 301-496-7157  
 Email: libin@helix.nih.gov  
 DNA Sequencing and analyses by National Institutes of Health  
 Intramural Sequencing Center (NISC).  
 Plate: 20 row: 9 column: 10  
 Seq primer: M13Rpl reverse primer (AB1).  
 Location/Qualifiers  
 1..374  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="NHTRC\_c20g10"  
 /sex="Female"  
 /tissue\_type="Bone"  
 /cell\_type="Trabecular Bone Cells"  
 /lab\_host="SURE"  
 /clone\_lib="Normal Human Trabecular Bone Cells"  
 /note="Organ: Hip; Vector: pBlueScript; Site\_1: EcoRI;  
 Library constructed by Dr. Marian Young and Dr. Pamela  
 Gehron Robey (NIDCR)"  
 BASE COUNT 55 a 138 c 124 g 57 t  
 ORIGIN

Query Match 37.0%; Score 137.8; DB 9; Length 374;  
 Best Local Similarity 95.3%; Pred. No. 4.2e-22;  
 Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 98 GTTCAGACCTGGCCCGCAGATGCTTCGGGAACCTGCAGAAACCAACCGCGCTGCAGG 157  
 DB 103 GCTCAGACCTGGCCCGCAGATGCTTCGGGAACCTGCAGAAACCAACCGCGCTGCAGG 162  
 QY 158 ACGTGGCGGACTGGCTGGCGAGCAGCTCAGGAGATCACTTCCTGAAAAACCGTGA 217  
 DB 163 ACGTGGCGGAGCTGCTGGCGAGCAGCTCAGGAGATCACTTCCTGAAAAACCGTGA 222  
 QY 218 TGGAGTGTGACGCGTGCAGCGCGCAGCCG 246  
 DB 223 TGGAGTGTGACGCGTGCAGCGCGATGCAGCAG 251

RESULT 11 BG900556 375 bp mRNA linear EST 06-NOV-2001  
 LOCUS HOA42-1-C6 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,  
 DEFINITION mRNA sequence.

ACCESSION BG900556 GI:14310805  
VERSION BG900556.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 375)  
Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,  
Sache,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and  
Lark,M.W.  
TITLE Identification and initial characterization of 5000 expressed  
sequenced tags (ESTs) each from adult human normal and  
osteochondr. cartil. 9 (7), 641-653 (2001)  
JOURNAL 21482651  
MEDLINE 11597177  
PUBMED 11597177  
COMMENT Contact: Sanjay Kumar  
UM2109  
GlaxoSmithKline  
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA  
Tel: 610-270-7245  
Fax: 610-270-5598  
Email: sanjay.kumar-1@sk.com  
Seq primer: 17.  
Location/Qualifiers  
1..375  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/issue\_type="cartilage"  
/lab\_host="E.coli DH10 B"  
/clone\_lib="HOA (Human Osteochondr. Cartilage)"  
/note="Vector: pSPORT 1; Site\_1: SalI; Site\_2: NotI;  
Directional"  
BASE COUNT 67 a 130 c 120 g 58 t  
ORIGIN  
Query Match 37.0%; Score 137.8; DB 12; Length 375;  
Best Local Similarity 95.3%; Pred. No. 4.2e-22;  
Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 98 GTTCAGACCTGGGCGCCGAGATGCTTGGGAACTGAGGAAACCAACGCGCGCTGCAG 157  
DB 177 GCTCAGACCTGGGCGCCGAGATGCTTGGGAACTGAGGAAACCAACGCGCGCTGCAG 236  
QY 158 ACGTGGGAGCTGGTGGCGGAGCAGGTGAGGAGATCAGTCTCTGAAAAACACGGTGA 217  
DB 237 ACGTGGGAGCTGGTGGCGGAGCAGGTGAGGAGATCAGTCTCTGAAAAACACGGTGA 296  
QY 218 TGAAGTGTGACCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 246  
DB 297 TGAAGTGTGACCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 325  
RESULT 12  
LOCUS BG898027 379 bp mRNA linear EST 06-NOV-2001  
DEFINITION HOA20-1-H3 HOA (Human Osteochondr. Cartilage) Homo sapiens cDNA,  
mRNA sequence.  
ACCESSION BG898027  
VERSION BG898027.1 GI:14308276  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 379)  
Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,  
Sache,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and  
Lark,M.W.  
TITLE Identification and initial characterization of 5000 expressed  
sequenced tags (ESTs) each from adult human normal and

JOURNAL 21482651  
MEDLINE 11597177  
PUBMED 11597177  
COMMENT Contact: Sanjay Kumar  
UM2109  
GlaxoSmithKline  
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA  
Tel: 610-270-7245  
Fax: 610-270-5598  
Email: sanjay.kumar-1@sk.com  
Seq primer: 17.  
Location/Qualifiers  
1..379  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/issue\_type="cartilage"  
/lab\_host="E.coli DH10 B"  
/clone\_lib="HOA (Human Osteochondr. Cartilage)"  
/note="Vector: pSPORT 1; Site\_1: SalI; Site\_2: NotI;  
Directional"  
BASE COUNT 67 a 134 c 119 g 59 t  
ORIGIN  
Query Match 37.0%; Score 137.8; DB 12; Length 379;  
Best Local Similarity 95.3%; Pred. No. 4.2e-22;  
Matches 142; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 98 GTTCAGACCTGGGCGCCGAGATGCTTGGGAACTGAGGAAACCAACGCGCGCTGCAG 157  
DB 181 GCTCAGACCTGGGCGCCGAGATGCTTGGGAACTGAGGAAACCAACGCGCGCTGCAG 240  
QY 158 ACGTGGGAGCTGGTGGCGGAGCAGGTGAGGAGATCAGTCTCTGAAAAACACGGTGA 217  
DB 241 ACGTGGGAGCTGGTGGCGGAGCAGGTGAGGAGATCAGTCTCTGAAAAACACGGTGA 300  
QY 218 TGAAGTGTGACCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 246  
DB 301 TGAAGTGTGACCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 329  
RESULT 13  
LOCUS BG896332 380 bp mRNA linear EST 06-NOV-2001  
DEFINITION HOA29-1-D12 HOA (Human Osteochondr. Cartilage) Homo sapiens cDNA,  
mRNA sequence.  
ACCESSION BG896332  
VERSION BG896332.1 GI:14306573  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 380)  
Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,  
Sache,G., Mul,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and  
Lark,M.W.  
TITLE Identification and initial characterization of 5000 expressed  
sequenced tags (ESTs) each from adult human normal and  
osteochondr. cartil. 9 (7), 641-653 (2001)  
JOURNAL 21482651  
MEDLINE 11597177  
PUBMED 11597177  
COMMENT Contact: Sanjay Kumar  
UM2109  
GlaxoSmithKline  
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA  
Tel: 610-270-7245  
Fax: 610-270-5598  
Email: sanjay.kumar-1@sk.com  
Seq primer: 17.  
Location/Qualifiers

	Best Local Similarity	95.3%	Pred No.	4.2e-22;	Mismatches	7;	Indels	0;	Gaps	0;
OY	Matches	142;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Db	191	GCTCAGACTGTGGGCGCCGACATGCTTGGGAACTGCAGAAAACCAACGCCGCGTCGAGG	250							
OY	158	ACGTGCGGGAACGTGCTGCCGACGACAGTCAAGGAGATCACGTTCTGTAAAAACAACGCTGA	217							
Db	251	ACGTCGCGGAGCTGCTGTCGGCAGCAGAGGTCAAGGATCAAGTCTTGAAAAACAACGCTGA	310							
OY	218	TGAGTGTGACGCGTGCAGGCGCCGACGCG	246							
Db	311	TGAGTGTGACGCGTGCAGGATGACAGCAG	339							
<hr/>										
RESULT_15										
LOCUS	BG898382	383 bp	mRNA	linear	EST	06-NOV-2001				
DEFINITION	HOA8-1-F4 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,									
ACCESSION	BG898382									
VERSION	BG898382									
KEYWORDS	EST.									
SOURCE	Homo sapiens (human)									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.									
	1 (bases 1 to 383)									
	Kumar,S., Connor,J.R., Dodde,R.A., Halsey,W., Van Horn,M., Mao,J.,									
	Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,U.C., Gowen,M. and									
	Lark,M.W.									
	Identification and initial characterization of 5000 expressed									
	sequenced tags (ESTs) each from adult human normal and									
	osteoarthritic cartilage cDNA libraries									
JOURNAL	Osteoarthr. Cartil. 9 (7), 641-653 (2001)									
MEDLINE	21482651									
PUBMED	11597177									
COMMENT	Contact: Sanjay Kumar									
	UW2109									
	GlaxoSmithKline									
	709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA									
	Tel: 610-270-7245									
	Fax: 610-270-5598									
	Email: sanjay.kumar-1@sk.com									
	Seq primer: T7.									
<hr/>										
FEATURES	Location/Qualifiers									
source	1..383									
	/organism="Homo sapiens"									
	/mol_type="mRNA"									
	/db_xref="taxon:9606"									
	/tissue_type="cartilage"									
	/lab_host="E.coli DH10 B"									
	/clone_lib="HOA (Human Osteoarthritic Cartilage)"									
	/note="Vector: pSPORT I, Site_1: SalI; Site_2: NotI;									
	directional"									
<hr/>										
BASE COUNT	67 a	137 c	120 g	59 t						
ORIGIN										
Query Match	37.0%;	Score 137.8;	DB 12;	Length 383;						
Best local Similarity	95.3%;	Pred. No. 4.2e-22;								
Matches 142;	Conservative	0;	Mismatches	7;	Indels	0;	Gaps	0;		
OY	98	GTTTCAGACCTGGGCCCCGACAGTGTCTTGGGAACTGCAGAAAACCAACGCCGCGTCGAGG	157							
Db	184	GCTCAGACTGTGGGCGCCGACATGCTTGGGAACTGCAGAAAACCAACGCCGCGTCGAGG	243							
OY	158	ACGTGCGGGAACGTGCTGCCGACGACAGTCAAGGAGATCACGTTCTGTAAAAACAACGCTGA	217							
Db	244	ACGTGCGGGAACGTGCTGCCGACGACAGTCAAGGAGATCACGTTCTGTAAAAACAACGCTGA	303							
OY	218	TGAGTGTGACGCGTGCAGGCGCCGACGCG	246							

Db 304 TGGAGTGTGACGCGGTGCGGATGCACAG 332

Search completed: August 26, 2003, 15:34:29  
Job time : 1434 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 26, 2003, 16:06:34 ; Search time 59 Seconds  
(without alignments)  
2001.569 Million cell updates/sec

Title: US-09-696-872-24  
Perfect score: 672  
Sequence: 1 aagcttaccatgggaagta.....aattccatcgcacccgcag 372

Scoring table:  
BIOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues  
Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p.model -DEV=rlp  
-Q=/cgnr\_1/USPTO.spool/US09696872/runat\_26082003\_150541\_8713/app\_query.fasta\_1.519  
-DB=A.Geneseq.19Jun03 -QPMT=fastran -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonsum2 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=pico -NOEM=ext -HEAPSIZE=500 -MUTLEN=0 -MAXLEN=200000000  
-USER=US09696872.QCGN\_1.1.117.rfunat\_26082003\_150541\_8713 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEODDERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
A.Geneseq.19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	579	86.2	109	21	AA44963	KDEL receptor inh1
2	459	68.3	109	21	AA44966	KDEL receptor inh1
3	436.5	65.0	115	21	AA44959	KDEL receptor inh1
4	429	63.8	109	21	AA44965	KDEL receptor inh1
5	429	63.8	134	21	AA44967	KDEL receptor inh1
6	422.5	62.9	115	21	AA44958	KDEL receptor inh1
7	379.5	56.5	109	21	AA44961	KDEL receptor inh1
8	353.5	52.6	109	21	AA44961	KDEL receptor inh1
9	345.5	51.4	105	21	AA44960	KDEL receptor inh1
10	304.5	45.3	90	21	AA44964	KDEL receptor inh1
11	262.5	39.1	242	21	AA800040	Human COMP/TSP-1 c
12	262.5	39.1	300	21	AA800041	Human COMP/TSP-2 c
13	262.5	39.1	757	21	AA800044	Human cartilage ol
14	262.5	39.1	757	23	AB005594	Breast cancer-asso
15	262.5	39.1	757	24	AB847420	Breast cancer asso
16	262.5	39.1	757	24	AB056605	Lung cancer-associ
17	262.5	39.1	758	23	AB882286	Human thrombospond
18	256.5	38.2	130	22	AA893197	Human polypeptide,
19	241	35.9	46	21	AA44947	Human cartilage ol
20	211	31.4	46	21	AA44946	Rat cartilage olig
21	153	22.8	1357	22	AA879715	Human protein SEQ
22	146.5	21.8	889	15	AA856248	Xenopus thrombospo
23	146	21.7	35	15	AA849650	Sequence of hinge
24	146	21.7	35	15	AA849650	Camel 2-chain Ig h
25	146	21.7	54	15	AA849722	Sequence of a pept
26	146	21.7	54	15	AA849536	Camel Ig 2-heavy c
27	145.5	21.7	776	22	AA879078	Human protein SEQ
28	145.5	21.7	961	15	AA856249	Human thrombospond
29	145.5	21.7	961	15	AA856249	Human thrombospond
30	144	21.4	24	21	AA850036	Target peptide #6
31	144	21.4	24	23	AA018488	Human antibody 11b
32	142.5	21.2	622	22	AA880062	Human protein SEQ
33	138	20.5	74	18	AA820047	Human TSP4 pentame
34	133	19.8	272	10	AA893560	Plasmodium berghei
35	129.5	19.3	163	23	AA814306	Sugarcanne proline
36	126.5	18.8	230	23	AA814307	Sugarcanne proline
37	126	18.8	329	23	AA814305	Sugarcanne proline
38	124	18.5	22	23	AA880744	Linker peptide #2
39	123.5	18.4	1247	22	AA827733	Human full-length
40	123.5	18.4	1247	22	AA878731	Human protein SEQ
41	123	18.3	814	22	ABG04441	Novel human diagno
42	121	18.0	46	21	AA44950	Human thrombospond
43	121	18.0	667	19	AA848760	BOP1 protein. Mus
44	118	17.6	378	12	AA814160	PRP 378. Trilicun
45	117	17.4	46	21	AA44951	Xenopus thrombospo

## ALIGNMENTS

RESULT 1  
ID AA44963 standard; Protein: 109 AA.  
XX AA44963;  
XX 23-MAY-2000 (first entry)  
XX KDEL receptor inhibitor protein-6.  
XX KDEL receptor inhibitor; heat shock protein; immune response;  
XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;  
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
XX infectious disease; allergy; autoimmune disease.  
XX Chimeric - Adenovirus E3.  
XX Chimeric - Homo sapiens.  
OS Chimeric - Camelus sp.

XX Key Location/Qualifiers  
FH 1..20  
FT /label= Signal\_peptide  
FT /note= "Derived from adenovirus E3"  
FT Domain 30..75  
FT /note= "Human COMP pentamerisation domain"  
FT 76..99  
FT /note= "Camel IgG linker domain"  
PN WO200006729-A1.  
PD 10-FEB-2000.  
XX 28-JUL-1999; 99WO-US17147.  
XX 29-JUL-1998; 98US-0124671.  
XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX Rothman JE, Mayhew M, Hoe MH;  
XX  
XX MPI; 2000-195296/17.  
XX N-PSDB; AAZ50497.  
XX  
XX Inhibitors of the KDEL receptor which comprises an oligomerization  
FT domain useful for promoting secretion of proteins which are normally  
PT retained within the cell -  
XX  
XX Disclosure; Fig 6; 87pp; English.  
XX  
XX The patent discloses the use of KDEL receptor inhibitor to promote  
CC secretion of proteins that are normally retained within the cell such as  
CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
CC shock proteins more accessible to the immune system and improves immune  
CC response to a target antigen. The inhibitor protein comprises several  
CC subunits where each subunit comprises an oligomerisation domain and has  
CC at its carboxy terminus a region which binds to a KDEL receptor. The  
CC target antigen may be associated with diseases including neoplasia such  
CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
CC infectious diseases, allergy or autoimmune diseases. The present  
CC sequence is KDEL receptor inhibitor comprising regions encoding a  
CC cleavable signal peptide; the oligomerisation domain of human cartilage  
CC oligomeric matrix protein (COMP) pentamerisation domain; a camel IgG  
CC linker domain and the carboxy-terminal sequence KDEL. The subsequence  
CC GDCC is an alteration of rat COMP which provides increased stability via  
CC disulphide bonds.  
XX  
XX SQ Sequence 109 AA;  
XX  
XX Alignment Scores:  
XX Pred. No.: 1,396-47 Length: 109  
XX Score: 579.00 Matches: 108  
XX Percent Similarity: 100.00% Conservative: 0  
XX Best Local Similarity: 100.00% Mismatches: 0  
XX Query Match: 86.16% Indels: 0  
XX DB: 21 Gaps: 0  
XX  
XX US-09-696-872-24 (1-372) x AAIV44963 (1-109)  
XX  
XX 16 AGTACATGATTTTGGCTGCTGCGCTTGCGGAGTCTGCGAGCGCTGCCAAAAAGGA 75  
XX 2 ArgTYrMetIleLeuGLyLeuLeuAlaLeuAlaValCysSerAlaAlaIshysIysGLy 21  
XX 76 TCCAGCCTGGGTGAGAGACTGTTGTTCAAGCTCGGCGCCGCGAGATCTTCGGGAATGCA 135  
XX 22 SerSerLeuGLyGLyASPcysCysSerAspLeuGLyProGlnMetLeuArgGluLeuGln 41  
XX 136 GAACCAACGGGGGCGTGGAGAGCTGCGGAGCTGGCTGGCGGACAGTCAAGGAGATC 195  
XX 42 GlnThrAsnAlaAlaLeuGlnAspValArgAspTrpLeuArgGlnGlnValArgGluIle 61

QY 196 ACGTTCCTGAAAAACACCGTGATGAGTGTGACCGCTGCGGGCCGACCCGACCCGAAA 255  
DB 62 ThreLeuIysAsnThrValMetGluCysAspAlaCysGlyProGlnProGlnProlys 81  
QY 256 CCGCAGCCCGGAGCGGACCGCCGACCCGAAACCGAAACCGGAAACCGGAAGGTTCC 315  
DB 82 ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnGlyThr 101  
QY 316 GGATCATCAGAAAAAGATGAGTTG 339  
DB 102 GlySerSerGluIysAspGluLeu 109  
XX  
XX RESULT 2  
XX ID AAIV44966 standard; Protein; 109 AA.  
XX  
XX AAIV44966;  
XX  
XX 23-MAY-2000 (first entry)  
XX  
XX KDEL receptor inhibitor protein-9.  
XX  
XX KDEL receptor inhibitor; heat shock protein; immune response;  
XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukemia;  
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
XX infectious disease; allergy; autoimmune disease.  
XX  
XX Chimeric - Adenovirus E3.  
XX Chimeric - Homo sapiens.  
XX Chimeric - Camelus sp.  
XX  
XX Key Location/Qualifiers  
FH 1..20  
FT /label= Signal\_peptide  
FT /note= "Derived from adenovirus E3"  
FT 30..75  
FT /note= "Human TSP4 trimerisation domain"  
FT 76..99  
FT Domain  
FT /note= "Camel IgG linker domain"  
XX  
XX WO200006729-A1.  
XX  
XX 10-FEB-2000.  
XX  
XX 28-JUL-1999; 99WO-US17147.  
XX  
XX 29-JUL-1998; 98US-0124671.  
XX  
XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX  
XX Rothman JE, Mayhew M, Hoe MH;  
XX  
XX MPI; 2000-195296/17.  
XX N-PSDB; AAZ50500.  
XX  
XX Inhibitors of the KDEL receptor which comprises an oligomerization  
FT domain useful for promoting secretion of proteins which are normally  
PT retained within the cell -  
XX  
XX Disclosure; Fig 9; 87pp; English.  
XX  
XX The patent discloses the use of KDEL receptor inhibitor to promote  
CC secretion of proteins that are normally retained within the cell such as  
CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
CC shock proteins more accessible to the immune system and improves immune  
CC response to a target antigen. The inhibitor protein comprises several  
CC subunits where each subunit comprises an oligomerisation domain and has  
CC at its carboxy terminus a region which binds to a KDEL receptor. The  
CC target antigen may be associated with diseases including neoplasia such  
CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
CC astrocytoma, with defective tumour suppressor genes, oncogenes,

CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence is KDEL receptor inhibitor comprising regions encoding a  
 CC cleavable signal peptide; the oligomerization domain of human  
 CC thrombospondin 4 (TSP4) trimerisation domain; a camel IgG linker domain  
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an  
 CC alteration of rat cartilage oligomeric matrix protein which provides  
 CC increased stability via disulphide bonds.

XX Sequence 109 AA;

#### Alignment Scores:

Pred. No.: 5.37e-36 Length: 109  
 Score: 459.00 Matches: 84  
 Percent Similarity: 86.11% Conservative: 9  
 Best Local Similarity: 77.78% Mismatches: 15  
 Query Match: 68.30% Indels: 0  
 DB: 21 Gaps: 0

US-09-696-872-24 (1-372) x AAY44966 (1-109)

QY 16 AGTATCATGATTAGGCTTGCGCCCTTGCGGAGTCGAGGGCTGCCAAAAAGGA 75  
 DB 2 ArgTyrMetIleLeuGlyLeuLeuAlaLeuAlaValCysSerAlaAlaIleuGly 21  
 QY 76 TCCAGCCTGGGTGAGACCTGTTGTTCAGACCTGGGCGCCGAGATGCTTGGGAATCGAG 135  
 DB 22 SerSerLeuGlyIleuGlyAspCysGlyGlyAspPheAsnArgGlnPheLeuGlyGlnMetThr 41  
 QY 136 GAAACCAACGGCGGCTGTCAGACGTCGGGAGATGCTGCGGACGACGTCAGAGATC 195  
 DB 42 GlnLeuAsnGlnLeuLeuGlyIleValIleAspLeuLeuArgGlnGlnValIleGlyIleThr 61  
 QY 196 ACGTTCCTGTAACAGCGGTATGAGGTGACCGCGTGGCGGCGCGGCGGACCGGAAA 255  
 DB 62 SerPheLeuAsnGlnThrIleAlaGlnCysGlnIleValCysGlyProGlnProGlnProlys 81  
 QY 256 CCGCAGCGCGACCGCGACCGCGACCGGAAACCGGACCGGAAACCGGAAAGTTAC 315  
 DB 82 ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 101  
 QY 316 GGATCATCAAAAAAGATGAGTTG 339  
 DB 102 GlySerSerGlnIleAspGlnLeu 109

#### RESULT 3

AAY44959 standard; Protein, 115 AA.

XX AAY44959;

DT 23-MAY-2000 (first entry)

XX KDEL receptor inhibitor protein-2.

XX KDEL receptor inhibitor; heat shock protein; immune response;  
 KM oligomerization domain; neoplasia; sarcoma; lymphoma; leukaemia;  
 KM melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
 KW infectious disease; allergy; autoimmune disease.

XX Chimeric - Mus sp.

OS Chimeric - Camelus sp.

OS Chimeric - Rattus sp.

XX Key Location/Qualifiers

FT 1..20 /label= "Signal peptide"

FT 26..30 /note= "Derived from mouse Bip"

FT /note= "Altered subsequence from rat COMP"

FT /note= "Rat COMP pentamerisation domain"

FT 82..105 /note= "Camel IgG linker domain"

XX WO200006729-A1.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17147.

XX 29-JUL-1998; 98US-0124671.

XX (SLOK) SLOAN KETTERING INST CANCER RES.

XX Rochman JE, Mayhew M, Hoe MH;

XX WPI; 2000-195296/17.

XX N-PSDB; AAZ50493.

XX Inhibitors of the KDEL receptor which comprises an oligomerization

XX retained within the cell

XX Disclosure; Fig 2; 87pp; English.

XX The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerization domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence is KDEL receptor inhibitor protein comprising regions including  
 CC a cleavable signal peptide; the oligomerization domain from rat  
 CC cartilage oligomeric matrix protein (COMP); a camel IgG linker domain  
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an  
 CC alteration of rat COMP which provides increased stability via disulphide  
 CC bonds.

XX Sequence 115 AA;

#### Alignment Scores:

Pred. No.: 8.06e-34 Length: 115  
 Score: 436.50 Matches: 89  
 Percent Similarity: 78.69% Conservative: 7  
 Best Local Similarity: 72.95% Mismatches: 19  
 Query Match: 64.96% Indels: 4  
 DB: 21 Gaps: 4

US-09-696-872-24 (1-372) x AAY44959 (1-115)

QY 10 ATGGAGATGATGATTGTTA-----GGCTTGCTCGCCCTTGCGGAGTCGAGCGCT 63  
 DB 1 MetClySerPheThrValIleAlaIleAlaLeuLeuLeuLeuGlyAlaVal----- 17  
 QY 64 GCCAAAAAGATTCACGCTGGGTGAGACTGTTGTTCAGACCTGGGCGCGACAGTCTT 123  
 DB 18 ArgIleGlnIleGlySerLeuGlyIleAspCys-----ProGlnMetLeu 33  
 QY 124 CCGGAATCGAGAAACCAACCGCGGCTGTCAGACCTGGGAGACTGCTGGGACACAG 183  
 DB 34 ArgGlnLeuGlnGlnGlnIleAlaIleAlaLeuGlnAspValArgGlnLeuLeuArgGln 53  
 QY 184 GTCAGGAGATCACTGCTGAACCAACCGGATGAGTGTGACGCGTCCGG----- 237  
 DB 54 ValIleGlnIleThrPheLeuIleAsnThrValMetClyCysAspAlaCysGlyMetGln 73  
 QY 238 -----CCGACCGCGACCGGAAACCGGACCGGACCGGACCGGACG 273  
 DB 74 ProAlaArgThrProGlyThrSerProGlnProGlnProGlnProGlnProGln 93

QY	27	CCGACGCCGAAACCGAGCCGAAACCGAGACCGGATCATGAGAAAAAT	333
Db	94	ProGlnProIleProGlnProIleProIleProIleProIleGlySerSerGluLysasp	1133
QY	334	GAGTTG 339	
Db	114	GluLeu 115	
RESULT 4			
AAV44965	ID	AAV44965 standard; Protein; 109 AA.	
AAV44965;	AC		
23-MAY-2000	DT	(first entry)	
KDEL receptor inhibitor protein-8.	XX		
KDEL receptor inhibitor; heat shock protein; immune response;	XX		
oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;	XX		
melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;	XX		
infectious disease; allergy; autoimmune disease.	XX		
Chimeric - Adenovirus E3.	OS		
Chimeric - Homo sapiens.	OS		
Chimeric - Camelus sp.	OS		
Key	XX	Location/Qualifiers	
Peptide	FT	1..20	
/label= Signal peptide	FT		
/note= "Derived from adenovirus E3"	FT		
Domain	FT	30..75	
/note= "Human TSP 3 trimerisation domain"	FT		
Domain	FT	76..99	
/note= "Camel Igg linker domain"	FT		
W0200006729-A1.	PN		
10-FEB-2000.	PD		
28-JUL-1999;	PF	99WO-US17147.	
29-JUL-1998;	PR	98US-0124671.	
(SLOK ) SLOAN KETTERING INST CANCER RES.	XX		
Rothman JE, Mayhew M, Hoe MH;	XX		
WPI; 2000-195296/17.	XX		
N-PSDB; AAZ50499.	XX		
Inhibitors of the KDEL receptor which comprises an oligomerization	PT		
domain useful for promoting secretion of proteins which are normally	PT		
retained within the cell	PT		
disclosure; Fig 8; 87pp; English.	PS		
The patent discloses the use of KDEL receptor inhibitor to promote	CC		
secretion of proteins that are normally retained within the cell such as	CC		
heat shock proteins by inhibiting KDEL receptor-mediated return of	CC		
protein complexes to endoplasmic reticulum. This makes the secreted heat	CC		
shock proteins more accessible to the immune system and improves immune	CC		
response to a target antigen. The inhibitor protein comprises several	CC		
subunits where each subunit comprises an oligomerisation domain and has	CC		
at its carboxy terminus a region which binds to a KDEL receptor. The	CC		
target antigen may be associated with diseases including neoplasia such	CC		
as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and	CC		
astrocytoma, with defective tumour suppressor genes, oncogenes,	CC		
infectious diseases, allergy or autoimmune diseases. The present	CC		
sequence is: KDEL receptor inhibitor comprising regions encoding a	CC		
cleavable signal peptide; the oligomerisation domain of human	CC		
thrombospondin 3 (TSP3) trimerisation domain; a camel Igg linker domain	CC		
and the carboxy-terminal sequence KDEL. The subsequent GDCC is an	CC		

CC	alteration of rat cartilage oligomeric matrix protein which provides
XX	increased stability via disulphide bonds.
XX	Sequence 109 AA,
Alignment Scores:	
Pred. No.:	4,24e-33
Score:	429.00
Percent Similarity:	84.25%
Best Local Similarity:	72.22%
Query Match:	63.84%
DB:	21
Gaps:	0
US-09-696-872-24 (1-372) x AAY44965 (1-109)	
QY	16 AGGTACATATTTTAGGCTTGCTCGCCCTTGCGGCAGCTTCGACGCTGCCAAAAAGCA 75
DB	2 ArgTylMetElleLeuGlyneuLeuAlaLeuAlaValCysSerAlaAlaLysGly 21
QY	76 TTCAGCGCTGGTGGAGACCTGTGTTTCAGACCTGGGCGCCGACAGTCTTCGGGACTGCAG 135
DB	22 SerSerLeuGlyGlyAspCysCysGlyIuGluGlnThrLysAlaLeuValThnGlnLeuThr 41
QY	136 GAAACCAACGCGCGCGCTGCAGACCTGCGGGAATGCTGCTCGCGCAGCAGTTCAGGAGATC 195
DB	42 LeuPheAsnGlnIleLeuValGluIuLeuArgAspIleArgAspGlnValLysGluMet 61
QY	196 AGGTCTCTAAAAACACGGGTGATGGAGGTGACGGGTGCGGGCGCGACCGCACCGGAA 255
DB	62 SerLeuIleArgAsnThrIleMetCylucysGlnValCysGlyProGlnProGlnProLys 81
QY	256 CCGCAGCCGCGACCGCGACCGCGACCGCGAAACCGGACCGGAAACCGGAACGGAGTACC 315
DB	82 ProGlnProGlnProGlnProGlnProLysProGlnProLysProGlnProLysProGlnGlyThr 101
QY	316 GGATCATCTCAAAAAGATGATGTTG 339
DB	102 GlySerSerGluLysAspGluLeu 109
RESULT 5	
ID	AAY44967 standard; Protein; 134 AA.
XX	AAY44967;
XX	23-MAY-2000 (first entry)
XX	DE
XX	KDEL receptor inhibitor protein-10.
XX	KDEL receptor inhibitor; heat shock protein; immune response;
XX	oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;
XX	melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;
XX	infectious disease; allergy; autoimmune disease.
OS	Chimeric - Mus sp.
OS	Chimeric - Rattus sp.
OS	Chimeric - Camelus sp.
XX	Key
XX	Location/Qualifiers
FT	Peptide 0
FT	/label= 1..20
FT	/label= Signal_peptide
FT	/note= "Derived from mouse Bip"
FT	Region 23..32
FT	/label= Myc_tag_sequence
FT	Modified-site 36..38
FT	/note= "Asn is N-glycosylated"
FT	Domain 49..94
FT	/note= "Rat COMP pentamerisation domain"
FT	Domain 101..124
FT	/note= "Camel IgG linker domain"
XX	MO200006729-A1.

The patent discloses the use of KDEL receptor inhibitor to promote secretion of proteins that are normally retained within the cell such as heat shock proteins by inhibiting KDEL receptor-mediated return of protein complexes to endoplasmic reticulum. This makes the secreted heat shock proteins more accessible to the immune system and improves immune response to a target antigen. The inhibitor oligomer comprises several subunits where each subunit comprises an oligomethylation domain and has at its carboxy terminus a region which binds to a KDEL receptor. The target antigen may be associated with diseases including neoplasia such as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and astrocytoma, with defective tumour suppressor genes, oncogenes, infectious diseases, allergy or autoimmune diseases. The present sequence encodes KDEL receptor inhibitor comprising regions encoding a cleavable signal peptide; a myc-tag; an N-glycosylation sequence, the oligomethylation domain of rat cartilage oligomeric matrix protein (COMP), a camel IgG linker domain and the carboxy-terminal sequence KDEL. The subsequence GDCC is an alteration of rat COMP which provides increased stability via disulphide bonds.

US-09-696-872-24 (1-372) x AAY4967 (1-134,

Dd		97	ProGlyThrSerProGlInPrroGlInProllysProGiInProGiInProGiInProGiInProlys
Oy	. .	286	CcGGACCCGAACCgAGCAGCGATCATGTAAAAAAGAATTG 339           + + + + + Dd
		117	ProGIInPolySProGIInProGIInGLuglYThrgISerSergIultYAspgIutLeu 134           + + + + +
<b>RESULT 6</b>			
ID	AAY44958	standard; Protein; 115 AA.	
XX AC	AAY44958;		
DT DT	23-MAY-2000	(first entry)	
DE DE	KDEL receptor inhibitor protein-1.		
KM KM	KDEL receptor inhibitor; heat shock protein; immune response; oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia; melanoma; carcinoma; glioblastoma; astrocytoma; oncogene; infectious disease; allergy; autoimmune disease. KM XX		
OS OS	Chimeric - Mus sp.		
OS OS	Chimeric - Camelus sp.		
XX XX	Chimeric - Rattus sp.		
FH FH	Key	Location/Qualifiers	
FT FT	Peptide	1..20	
FT FT	/label= Signal peptide		
FT FT	/note= "Derived from mouse Bfp"		
Domain Domain		26..30	
FT FT	/note= "Subsequence from rat COMP"		
Domain Domain		30..71	
FT FT	/note= "Rat COMP pentamerisation domain"		
FT FT	82..105		
PEN PN	/note= "Camel IgG linker domain"		
XX XX	NQZ00006729-A1.		
XX XX	10-FEB-2000.		
PD PD	28-JUL-1999; 99WO-US17147.		
PR PR	29-JUL-1998; 98US-0124671.		
PA PA	(SLOK ) SLOAN KETTERING INST CANCER RES.		
XN XN	Rothman JE, Mayhew M, Hoe MH;		
DR DR	WPI, 2000-195296/17.		
DR DR	N-PsDB; AAZ50492.		
PT PT	Inhibitors of the KDEL receptor which comprises an oligomerization domain useful for promoting secretion of proteins which are normally retained within the cell -		
PS PS	Disclosure; Fig 1; 87pp; English.		
CC CC	The patent discloses the use of KDEL receptor inhibitor to promote secretion of proteins that are normally retained within the cell such as heat shock proteins by inhibiting KDEL receptor-mediated return of protein complexes to endoplasmic reticulum. This makes the secreted heat shock proteins more accessible to the immune system and improves immune response to a target antigen. The inhibitor protein comprises several subunits where each subunit comprises an oligomerisation domain and has at its carboxy terminus a region which binds to a KDEL receptor. The target antigen may be associated with diseases including neoplasia such as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and astrocytoma, with defective tumour suppressor genes, oncogenes, infectious diseases, allergy or autoimmune diseases. The present sequence is KDEL receptor inhibitor protein comprising regions including a cleavable signal peptide; the oligomerisation domain from rat cartilage oligomeric matrix protein; a camel igg linker domain and the carboxy-terminal sequence KDEL.		

XX	Sequence	115 AA:	
SQ	Alignment Scores:		
XX	Pred. No.:	1,81e-32	Length: 115
XX	Score:	422.50	Matches: 88
XX	Percent Similarity:	77.87%	Conservative: 7
XX	Best Local Similarity:	72.13%	Mismatches: 8
XX	Query Match:	62.87%	Indels: 19
DB:		21	Gaps: 4
US-09-696-872-24 (1-372) x AAY44958 (1-115)			
OY	10	ATGGGAGGTACGATGATTTA-----GGCTTGCTCGCCCTTGCGGAGCTTCGACGCGCT	63
DB	1	MetGlyLysPheThrValValAlaAlaAlaLeuLeuLeuGlyAlaVal-----	17
OY	64	GCCAAAAAGATCCAGCTCGGTGGAGACTGTTTCAGACCTGGGCGCCGACATGCTT	123
DB	18	ArgAlaGluGlySerSerLeuGlyGly-----AspLeuAlaProGlnMetLeu	33
OY	124	CGGGAATCGAGAAACCAACCGCGCGCTGCAGACGTGCGGAACTGGCTGCAGCAG	183
DB	34	ArgGluLeuGlnGlnThrPheAlaAlaLeuGlnAspValArgGluLeuArgGlnGln	53
OY	184	GTCAGGAGATCAGCTTCTGAAAAACAAGGTGAGTGTGACGCGCTGCGG-----	237
DB	54	ValLysGluIleThrPheLeuLysAsnThrValMetGluCysAspAlaCysGlyMetGln	73
OY	238	-----CCGCAGCGCGACCGGAAACCGCACCGCGACCGCGCAG	273
DB	74	ProAlaArgThrProGlyThrSerProGlnProGlnProGlnProGlnProGln	93
OY	274	CCGCAGCGGAAACCGCACCGGAAACCGGAAACCGGAAAGGTACCGATCATGAAAAAGAT	333
DB	94	ProGlnProLysProGlnProLysProGlnProGlnGlyThrGlySerSerGluLysAsp	113
OY	334	GAGTTG 339	
DB	114	GluLeu 115	
RESULT 7			
AAY44962			
ID	AAY44962 standard; Protein; 109 AA.		
XX	AAY44962;		
XX	AC		
XX	DT	23-MAY-2000 (first entry)	
XX	XX	KDEL receptor inhibitor protein-5.	
XX	XX	KDEL receptor inhibitor; heat shock protein; immune response;	
KW		oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;	
KW		melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;	
XX		infectious disease; allergy; autoimmune disease.	
XX		Chimeric - Mus sp.	
OS		Chimeric - Xenopus sp.	
OS		Chimeric - Camelus sp.	
OS		Chimeric - Rattus sp.	
XX		Key	
FT	FT	Peptide	1..20
FT		/label=	Signal peptide
FT		/note=	Derived from mouse B1p"
FT		Domain	26..30
FT		/note=	Altered subsequence from rat COMP"
FT		Domain	30..75
FT		/note=	"Xenopus thrombospondin 4 trimerisation domain"
FT		Domain	76..99
FT		/note=	"Camel IgG linker domain"
XX			
PN		W0200006729-A1.	

XX	Sequence	109 AA:	
PD	10-FEB-2000.		
XX	XX	28-JUL-1999; 99WO-US17147.	
XX	XX	29-JUL-1998; 98US-0124671.	
XX	XX	(SLOK ) SLOAN KETTERING INST CANCER RES.	
XX	XX	Roelman JE, Mayhew M, Hoe MH;	
XX	XX	WPI; 2000-195296/17.	
DR	N-PSDB; AAZ50496.		
PT	Inhibitors of the KDEL receptor which comprises an oligomerization		
PT	domain useful for promoting secretion of proteins which are normally		
PT	retained within the cell		
XX	PS	Disclosure; Fig 5; 87pp; English.	
XX	XX	The patent discloses the use of KDEL receptor inhibitor to promote	
CC	CC	secretion of proteins that are normally retained within the cell such as	
CC	CC	heat shock proteins by inhibiting KDEL receptor-mediated return of	
CC	CC	protein complexes to endoplasmic reticulum. This makes the secreted heat	
CC	CC	shock proteins more accessible to the immune system and improves immune	
CC	CC	response to a target antigen. The inhibitor protein comprises several	
CC	CC	subunits where each subunit comprises an oligomerisation domain and has	
CC	CC	at its carboxy terminus a region which binds to a KDEL receptor. The	
CC	CC	target antigen may be associated with diseases including neoplasia such	
CC	CC	as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and	
CC	CC	astrocytoma, with defective tumour suppressor genes, oncogenes,	
CC	CC	infectious diseases, allergy or autoimmune diseases. The present	
CC	CC	sequence is KDEL receptor inhibitor comprising regions encoding a	
CC	CC	cleavable signal peptide; the oligomerisation domain of Xenopus	
CC	CC	thrombospondin 4 (TSP4) trimerisation domain including an additional	
CC	CC	subsequence; a camel IgG linker domain and the carboxy-terminal sequence	
CC	CC	KDEL. The subsequence GDCC is an alteration of rat cartilage oligomeric	
XX	XX	matrix protein which provides increased stability via disulphide bonds.	
SQ	Sequence	109 AA;	
Alignment Scores:			
Pred. No.:	2.55e-28	Length: 109	
Score:	379.50	Matches: 70	
Percent Similarity:	77.68%	Conservative: 17	
Best Local Similarity:	62.50%	Mismatches: 20	
Query Match:	56.47%	Indels: 5	
DB:	21	Gaps: 2	
US-09-696-872-24 (1-372) x AAY44962 (1-109)			
OY	10	ATGGGAGGTACGATGATTTA-----GGCTTGCTCGCCCTTGCGGAGCTTCGACGCGCT	63
DB	1	MetGlyLysPheThrValValAlaAlaAlaLeuLeuLeuGlyAlaVal-----	17
OY	64	GCCAAAAAGATCCAGCTCGGTGGAGACTGTTTCAGACCTGGGCGCCGACATGCTT	123
DB	18	ArgAlaGluGlySerSerLeuGlyGlyAspCysGlySerValSerArgGlnLeuIle	37
OY	124	CGGGAATCGAGAAACCAACCGCGCGCTGCAGACCTGGGGAATGCTTCGCGCAG	183
DB	38	GlyGlnIleThrGlnMetLeuGlnMetLeuGlyGluLeuArgAspValMetArgGlnGln	57
OY	184	GTCAGGAGATCAGCTTCTGAAAAACAAGGTGAGTGTGACGCTGGGCGCGCGCAG	243
DB	58	ValLysGluThrMetPheLeuArgAsnThrIleAlaGluCysGlnAlaCysGlyProGln	77
OY	244	CCGCAGCGGAAACCGCACCGCGACCGCGACCGCGAAACCGGAAACCGGAA	303
DB	78	ProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProLysProGln	97
OY	304	CCGGAAGGTACCGGATCATGAAAAAGATGAGTTG 339	

Db 98 ProglutinyThryglyserGluLysaspGluLeu 109

RESULT 8  
ID AAY44961 standard; Protein; 109 AA.  
AC AAY44961;  
XX 23-MAY-2000 (first entry)  
XX  
XX KDEL receptor inhibitor protein-4.  
XX  
XX KDEL receptor inhibitor; heat shock protein; immune response;  
XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;  
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
XX infectious disease; allergy; autoimmune disease.  
XX  
XX Chimeric - Mus sp.  
XX Chimeric - Camelus sp.  
XX Chimeric - Rattus sp.  
XX  
XX Key Location/Qualifiers  
XX Peptide 1..20  
XX /label= Signal peptide  
XX /note= "Derived from mouse B1p"  
XX 26..30  
XX /note= "Altered subsequence from rat COMP"  
XX 30..75  
XX /note= "Mouse thrombospondin 3 trimerisation domain"  
XX 76..99  
XX /note= "Camel IgG linker domain"  
XX  
XX MO200006729-A1.  
XX  
XX 10-FEB-2000.  
XX  
XX 28-JUL-1999; 99WO-US17147.  
XX  
XX 29-JUL-1998; 98US-0124671.  
XX  
XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX  
XX Rothman JE, Mayhew M, Hoe MH;  
XX  
XX WPI: 2000-195296/17.  
XX N-PSDB; AAZ50495.  
XX  
XX Inhibitors of the KDEL receptor which comprises an oligomerization  
XX domain useful for promoting secretion of proteins which are normally  
XX retained within the cell  
XX  
XX  
XX Disclosure; Fig 4; 87pp; English.  
XX  
XX The patent discloses the use of KDEL receptor inhibitor to promote  
XX secretion of proteins that are normally retained within the cell such as  
XX heat shock proteins by inhibiting KDEL receptor-mediated return of  
XX protein complexes to endoplasmic reticulum. This makes the secreted heat  
XX shock proteins more accessible to the immune system and improves immune  
XX response to a target antigen. The inhibitor protein comprises several  
XX subunits where each subunit comprises an oligomerisation domain and has  
XX at its carboxy terminus a region which binds to a KDEL receptor. The  
XX target antigen may be associated with diseases including neoplasia such  
XX as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
XX astrocytoma, with defective tumour suppressor genes, oncogenes,  
XX infectious diseases, allergy or autoimmune diseases. The present  
XX sequence is KDEL receptor inhibitor comprising regions encoding a  
XX cleavable signal peptide; the oligomerisation domain including an additional  
XX thrombospondin 3 (TSP3) trimerisation domain including an additional  
XX subsequence; a camel IgG linker domain and the carboxy-terminal sequence  
XX KDEL. The subsequence GCCC is an alteration of rat cartilage oligomeric  
XX matrix protein which provides increased stability via disulphide bonds.  
XX  
XX Sequence 109 AA;

Alignment Scores:

Pred. No.:	8 26e-26	Length:	109
Score:	353.50	Matches:	66
Percent Similarity:	75.00%	Conservative:	18
Best Local Similarity:	58.93%	Mismatches:	23
Query Match:	52.60%	Indels:	5
DB:	21	Gaps:	2

US-09-696-872-24 (1-372) x AAY44961 (1-109)

QY 10 ATGGGAAGTACATGATTTTA-----GGCTTGCTCGCCCTTGCGGAGTCTGCAGCGCT 63  
Db 1 MetclyspherThrValAlaAlaAlaLeuLeuLeuLeuGlyAlaVal----- 17  
QY 64 GCCAAAAAAGATCCACCTCGGTGGAGACTGTTGTTCAACCTGGGCGCCGACATGCTT 123  
Db 18 ArgAlaGluGlySerSerLeuGlyGlyAspCysGlyGluGlnThrValAlaLeuVal 37  
QY 124 CGGGAACCTGACGAAACCAACGCGCGCTGCAGACGCGGAGCTGCTGCAGCAG 183  
Db 38 ThrGlnLeuThrLeuPheAsnGlnIleLeuValGluLeuArgAspAlaIleArgAspGln 57  
QY 184 GTCGAGGATCAACGTTCTGAAAAACACGCGTGAATGAGTGTGACGCGTCCGCGCAG 243  
Db 58 ValysGluMetSerLeuIleArgAsnThrIleMetGluCysGlnValCysGlyProGln 77  
QY 244 CCGGACCGCAACCGCAGCCGCGCAGCCGACCCGCAACCGCAACCGCAACCGGAA 303  
Db 78 ProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProLysProGln 97  
QY 304 CCGGAAGTACCGGATCATCAAGAAAGATGAGTGTG 339  
Db 98 ProglutinyThryglyserGluLysaspGluLeu 109

RESULT 9  
ID AAY44960 standard; Protein; 105 AA.  
AC AAY44960;  
XX  
XX 23-MAY-2000 (first entry)  
XX  
XX KDEL receptor inhibitor protein-3.  
XX  
XX KDEL receptor inhibitor; heat shock protein; immune response;  
XX oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;  
XX melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
XX infectious disease; allergy; autoimmune disease.  
XX  
XX Chimeric - Mus sp.  
XX Chimeric - Camelus sp.  
XX Chimeric - Rattus sp.  
XX  
XX Key Location/Qualifiers  
XX Peptide 1..20  
XX /label= Signal peptide  
XX /note= "Derived from mouse B1p"  
XX 26..30  
XX /note= "Altered subsequence from rat COMP"  
XX 30..71  
XX /note= "Mouse thrombospondin 3 trimerisation domain"  
XX 72..95  
XX /note= "Camel IgG linker domain"  
XX  
XX MO200006729-A1.  
XX  
XX 10-FEB-2000.  
XX  
XX 28-JUL-1999; 99WO-US17147.  
XX  
XX 29-JUL-1998; 98US-0124671.  
XX

PA (SLOK ) SLOAN KETTERING INST CANCER RES.  
 XX Rothman JE, Mayhew M, Hoe MH;  
 XX MPI: 2000-195296/17.  
 DR N-PSDB; AAZ50494.  
 XX  
 XX Inhibitors of the KDEL receptor which comprises an oligomerization  
 PT domain useful for promoting secretion of proteins which are normally  
 PT retained within the cell -  
 XX  
 XX Disclosure: Fig 3; 87pp; English.  
 XX  
 XX The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerization domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence is KDEL receptor inhibitor comprising regions encoding a  
 CC cleavable signal peptide; the oligomerisation domain of mouse  
 CC thrombospondin 3 (TSP3) trimerisation domain; a camel IgG linker domain  
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an  
 CC alteration of rat cartilage oligomeric matrix protein which provides  
 CC increased stability via disulphide bonds.  
 XX  
 XX Sequence 105 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 4,866-25 Length: 105  
 Score: 345.50 Matches: 67  
 Percent Similarity: 74.11% Conservative: 16  
 Best Local Similarity: 59.82% Mismatches: 20  
 Query Match: 51.41% Indels: 9  
 DB: 21 Gaps: 3  
 US-09-696-872-24 (1-372) x AAY44960 (1-105)  
 QY 10 ATGGAGAGTACATGATTTTA-----GGCTTGCTCGCCTTGGCAGTGTGACGCCT 63  
 DB 1 MetGlyLysPheThrValValAlaAlaLeuLeuLeuGlyAlaVal----- 17  
 QY 64 GCCAAAAAAGATCCAGCCTGGGTGAAGACTGTTGTTCAGACTGGGCCCGCAGATGCTT 123  
 DB 18 ArgAlaGluGlySerSerLeuGlyLysPheCysGlyAlaLeu-----Val 33  
 QY 124 CGGGAACTCGACGGAACCAACCGCGCGCTGCGAGACCTGCGGAGTGGCTGGCAGCAG 183  
 DB 34 ThrGlnLeuThrLeuPheAsnGlnLeuValGlnLeuAlaGlyAspAlaPheArgAspGln 53  
 QY 184 GTCAAGGAGATCACTCTCTGTAACCAACGAGTGAAGTGAAGCGTGGCGCGCAG 243  
 DB 54 VallysGluMetSerLeuLeuLeuArgThrIleMetGluCysGlnValCysGlyProGln 73  
 QY 244 CGCAGCCGGAACCGCAGCGCGCAGCGCGACCGCGAACCAGACCGGAACCGGAA 303  
 DB 74 ProGlnProLysPheProGlnProGlnProGlnProGlnProLysPheProGln 93  
 QY 304 CCGAAGGTACCGGATCATCAGAAAAAGATGAGTGG 339  
 DB 94 ProGlnGlyThrGlySerSerLeuLysAspGlnLeu 105  
 RESULT 10  
 ID AAY44964 standard; Protein; 90 AA.  
 XX AAY44964;  
 AC AAY44964;

XX 23-MAY-2000 (first entry)  
 DT  
 XX KDEL receptor inhibitor protein-7.  
 DE  
 XX  
 XX KDEL receptor inhibitor; heat shock protein; immune response;  
 KW oligomerisation domain; neoplasia; sarcoma; lymphoma; leukaemia;  
 KW melanoma; carcinoma; glioblastoma; astrocytoma; oncogene;  
 KW infectious disease; allergy; autoimmune disease.  
 XX  
 OS Chimeric - Adenovirus E3.  
 OS Chimeric - Homo. sapiens.  
 OS Chimeric - Camelus sp.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..20  
 FT /note= "Derived from adenovirus E3"  
 FT Domain 30..56  
 FT /note= "Human phospholamban pentamerisation domain"  
 FT Domain 57..80  
 FT /note= "Camel IgG linker domain"  
 XX  
 XX WO200006729-A1.  
 XX  
 XX 10-FEB-2000.  
 PD  
 XX  
 XX 28-JUL-1999; 99WO-US17147.  
 PF  
 XX 29-JUL-1998; 98US-0124671.  
 PR  
 XX  
 XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
 PA  
 PI Rothman JE, Mayhew M, Hoe MH;  
 XX MPI: 2000-195296/17.  
 DR N-PSDB; AAZ50498.  
 XX  
 XX Inhibitors of the KDEL receptor which comprises an oligomerization  
 PT domain useful for promoting secretion of proteins which are normally  
 PT retained within the cell -  
 XX  
 XX Disclosure: Fig 7; 87pp; English.  
 PS  
 XX The patent discloses the use of KDEL receptor inhibitor to promote  
 CC secretion of proteins that are normally retained within the cell such as  
 CC heat shock proteins by inhibiting KDEL receptor-mediated return of  
 CC protein complexes to endoplasmic reticulum. This makes the secreted heat  
 CC shock proteins more accessible to the immune system and improves immune  
 CC response to a target antigen. The inhibitor protein comprises several  
 CC subunits where each subunit comprises an oligomerisation domain and has  
 CC at its carboxy terminus a region which binds to a KDEL receptor. The  
 CC target antigen may be associated with diseases including neoplasia such  
 CC as sarcoma, lymphoma, leukemia, melanoma, carcinoma, glioblastoma and  
 CC astrocytoma, with defective tumour suppressor genes, oncogenes,  
 CC infectious diseases, allergy or autoimmune diseases. The present  
 CC sequence is KDEL receptor inhibitor comprising regions encoding a  
 CC cleavable signal peptide; the oligomerisation domain of human  
 CC phospholamban (PLB) pentamerisation domain; a camel IgG linker domain  
 CC and the carboxy-terminal sequence KDEL. The subsequence GDCC is an  
 CC alteration of rat cartilage oligomeric matrix protein which provides  
 CC increased stability via disulphide bonds.  
 XX  
 XX Sequence 90 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 4,326-21 Length: 90  
 Score: 304.50 Matches: 67  
 Percent Similarity: 60.18% Conservative: 16  
 Best Local Similarity: 59.29% Mismatches: 29  
 Query Match: 45.31% Indels: 2  
 DB: 21 Gaps: 2  
 US-09-696-872-24 (1-372) x AAY44964 (1-90)





**Nucleic acids encoding chimeric proteins such as cartilage oligomeric matrix protein (COMP)/thrombospondins (TSP)-1 and 2, useful for inhibiting angiogenesis and treating diseases such as cancer**

New nucleic acids are described which encode a protein comprising the second and third type 1 repeats of human TSP (thrombospondin)-1, but not the TSP (transforming growth factor)-beta activation region of human TSP-1. The nucleic acid of TSP (thrombospondin)-1 containing the second and third type-1 repeats and the COMP (cartilage oligomeric matrix protein) assembly sequence (COMP/TSP-1) was produced by PCR (polymerase chain reaction). Expression of COMP/TSP-1 caused inhibition of the growth of tumours in mice models. Thus the nucleic acids and proteins may be useful for treating angiogenesis related diseases such as cancer (by reducing the rate of growth and size of tumours), arthritis, psoriasis, diabetic retinopathy, corneal graft rejection, and glioma. They may also be used for treating human immunodeficiency virus (HIV) infection. Anti-angiogenic therapy has little toxicity, does not require the therapeutic agent to enter tumour cells or cross the blood-brain barrier, controls tumour growth independently of growth of tumour cell heterogeneity, and does not induce drug resistance.

SQ Sequence 300 AA;

Alignment Scores:	
Pred. No.:	5,93e-17
Score:	262.50
Percent Similarity:	75.61%
Best Local Similarity:	69.51%
Query Match:	39.06%
DB:	21
length:	300
Matches:	57
Conservative:	5
Mismatches:	9
Indels:	11
Gaps:	0

QY 214 GTGATGAGTGTGACGGCTGCGGCGCCGAG-----CCGACGCCG 252  
DB 66 ValMetGluCysAspAlaCysGlyMetGlnGlnSerValArgThrGlyLeuProSerVal 85  
QY 253 AAACCG 258  
DB 86 ArgPro 87  
RESULT 14  
ABU05594  
ID ABU05594 standard; Protein; 757 AA.  
XX AC ABU05594;  
XX 14-NOV-2002 (first entry)  
XX DE Breast cancer-associated protein 59.  
XX DE Breast cancer-associated protein 59.  
XX KW Breast cancer; breast cancer-associated gene sequence;  
XX KW drug development; pharmacogenetics; biosensor development.  
XX OS Unidentified.  
XX WO200259377-A2.  
XX PD 01-AUG-2002.  
XX PF 24-JAN-2002; 2002WO-US02242.  
XX PR 24-JAN-2001; 2001US-263965P.  
XX PR 02-FEB-2001; 2001US-265928P.  
XX PR 09-APR-2001; 2001US-282698P.  
XX PR 09-APR-2001; 2001US-0829472.  
XX PR 04-MAY-2001; 2001US-288590P.  
XX PR 29-MAY-2001; 2001US-294443P.  
XX PA (BOSB-) EOS BIOTECHNOLOGY INC.  
XX PI Mack DH, Gish KC, Afar D;  
XX DR WPI; 2002-583738/62.  
XX DR N-PSDB; ABT07751.  
XX PT Detecting a breast cancer-associated transcript in a patient's cell,  
XX PT useful for diagnosing breast cancer, comprises contacting a biological  
XX PT sample with a polynucleotide that selectively hybridizes with breast  
XX PT cancer nucleic acids  
XX PS  
XX PS Disclosure; Page 401; 414pp; English.  
XX CC The invention comprises a method of detecting a breast cancer-associated  
XX CC transcript in a cell from a patient. The method of the invention involves  
XX CC contacting a biological sample from the patient with a nucleotide that  
XX CC hybridizes to one of the 69 breast cancer-associated gene sequences shown  
XX CC in the specification. The method of the invention is useful in the  
XX CC diagnosis or prognosis of breast cancer, and for detecting genes that are  
XX CC up or down-regulated in breast cancer cells. Genes identified by the  
XX CC method of the invention can be used in diagnostic purposes and also as  
XX CC targets for screening for therapeutic compounds that modulate breast  
XX CC cancer (e.g. hormones or antibodies). Identification of genes that are  
XX CC over or under expressed in breast cancer can additionally provide high-  
XX CC resolution, high-sensitivity datasets which can be used in the areas of  
XX CC diagnostics, therapeutics, drug development, pharmacogenetics, protein  
XX CC structure and biosensor development. Amino acid sequences ABU0536 -  
XX CC ABU05604 represent the proteins encoded by the 69 breast cancer-  
XX CC associated genes of the invention.  
XX SQ Sequence 757 AA;

Best Local Similarity: 69.51% Mismatches: 9  
Query Match: 39.06% Indels: 11  
DB: 23 Gaps: 2  
US-09-696-872-24 (1-372) x ABU05594 (1-757)  
QY 34 TTGCTCGCCCTTGGCGAGTCTGCAGCGCTGCCAAAAAGATCCAGCTGGTGAGAC 93  
DB 10 LeuLeuThrLeuAlaAlaLeuGlyAlaSerGlyGlnGlyGlnSerProLeuGly----- 27  
QY 94 TGTGTTCAAGACTCGGAGCCCGGACATCTTCGGAATCTGCAGAAACCAACGCGCGCTG 153  
DB 28 -----SerAspLeuGlyProGlnMetLeuArgGlnLeuGlnGlnThrAsnAlaAlaLeu 45  
QY 154 CAGGACGTGGCGGAGCTGCTGCGGCGGACAGATCAAGGAGATCAAGTCTGAAAAACAGC 213  
DB 46 GlnAspValArgAspTrpLeuArgGlnGlnValArgGlnLeuThrPheLeuValAsnThr 65  
QY 214 GTGATGAGTGTGACGGCTGCGGCGCCGAG-----CCGACGCCG 252  
DB 66 ValMetGluCysAspAlaCysGlyMetGlnGlnSerValArgThrGlyLeuProSerVal 85  
QY 253 AAACCG 258  
DB 86 ArgPro 87  
RESULT 15  
ABR47420  
ID ABR47420 standard; Protein; 757 AA.  
XX AC ABR47420;  
XX DT 12-JUN-2003 (first entry)  
XX DE Breast cancer associated protein sequence SEQ ID NO:72.  
XX DE Human; breast cancer; cytostatic; gene therapy.  
XX KW Homo sapiens.  
XX OS  
XX WO2003004989-A2.  
XX PD 16-JAN-2003.  
XX PF 21-JUN-2002; 2002WO-US19669.  
XX PR 21-JUN-2001; 2001US-299887P.  
XX PR 18-JUN-2001; 2001US-301572P.  
XX PR 18-JUN-2001; 2001US-306501P.  
XX PR 25-SEP-2001; 2001US-325002P.  
XX PR 05-MAR-2002; 2002US-362585P.  
XX PR 14-MAY-2002; 2002US-380391P.  
XX PA (MILL-) MILLENIUM PHARM INC.  
XX PI Lillie J, Gannavarapu M, Glatc K, Hoerish S, Kamatkar S, Mertens M,  
XX PI Monahan JB, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE, Baer RC;  
XX PI Horrobaegy GN, Fuszta L, Meric F, Sahin A, Mills GB;  
XX DR WPI; 2003-210381/20.  
XX DR N-PSDB; ACC50112.  
XX PT Breast cancer diagnosis or treatment by comparing the level of  
XX PT expression of a marker in a patient sample with that in the control  
XX PT non-breast cancer sample  
XX PS  
XX PS Claim 1; SEQ ID 72; 128pp; English.  
XX CC The present invention describes a method for assessing whether a patient  
XX CC is afflicted with breast cancer. The method comprises comparing the level  
XX CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and  
XX CC ABR47386 to ABR47632) in a patient sample and the normal level of  
XX CC expression of the marker in a control non-breast cancer sample, where a



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 26, 2003, 16:19:24 ; Search time 15.5 Seconds  
(without alignments)  
2030.921 Million cell updates/sec

```
Title: US-09-696-872-24
Perfect score: 672
Sequence: 1 aagcttaccatcggaagtra.....aattccatattgcatctcgag 372
```

Scoring table:	
BLOSUM62	
Xgapop 10.0 ,	Xgapext 0.5
Ygapop 10.0 ,	Ygapext 0.5
Fgapop 6.0 ,	Fgapext 7.0
Delop 6.0 ,	Delext 7.0

Searched: 328717 seqs, 42310858 residues  
Total number of hits satisfying chosen parameters: 657434

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 s

```

```

Command line parameters:
-MODEL=Firemax_n2p_model -DEV=x1p
-D=/csgn2/1/USPO.spool/US095696872/runat_26082003_150544_8745/rap_query_fasta_1.519
-DB=Issued_Patents_AA -GMT=fasteran -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=51c -START=1 -END=1 -MATRIX=010sume2 -TRANS=numan40.cdi
-LIST=45 -POCALIGN=200 -THR SCORE=Pct -THR MAX=100 -THR MIN=0 -ALIGN=11
-MODE=LOCAL -OUTTYPE=pbio -NORM=ext HEADSIZE=500 -MINTEN=0 -MAXLEN=2000000000
-USER=US095696872_@CNC1_1_1_38_@rnatc_26082003_150544_8745 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESDENTRY -NEG_SCORES=0 -WAIT -DSBBLOCK=100 -LONGGLOG
-DEV_TIMESTEP=1210 -WARN_TIMESTEP=0 -THRAD=1 -XGAPEXT=0.5 -FCGAPD=6
-FCGAPEXT=7 -YGAPD=10 -YGAPEXT=0.5 -DLEXP=6 -DELEXT=7

```

```
Database : Issued Patents AA:*
1 : /cgn2_6/prodata/1/aa/5A.COMB.pep:*
2 : /cgn2_6/prodata/1/aa/5B.COMB.pep:*
3 : /cgn2_6/prodata/1/aa/6A.COMB.pep:*
4 : /cgn2_6/prodata/1/aa/6B.COMB.pep:*
5 : /cgn2_6/prodata/1/aa/PCUTS.COMB.pep:*
6 : /cgn2_6/prodata/1/aa/backfiles.pep:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	DB	ID	Description
1	579	86.2	109	3	US-09-124-671-23	Sequence 23, App
2	459	68.3	109	3	US-09-124-671-29	Sequence 29, App
3	436.5	65.0	115	3	US-09-124-671-15	Sequence 15, App
4	429	63.8	109	3	US-09-124-671-27	Sequence 27, App
5	429	63.8	134	3	US-02-124-671-34	Sequence 34, App
6	422.5	62.9	115	3	US-09-124-671-13	Sequence 13, App
7	379.5	56.5	109	3	US-09-124-671-21	Sequence 21, App
8	353.5	52.6	109	3	US-09-124-671-19	Sequence 19, App
9	345.5	51.4	105	3	US-09-124-671-17	Sequence 17, App
10	304.5	45.3	90	3	US-09-124-671-25	Sequence 25, App
11	241	35.9	46	3	US-02-124-671-2	Sequence 2, App
12	234.5	34.9	66	3	US-09-091-814-46	Sequence 46, App

13	228	3.9	64	3	US-09-091-814-45	Sequence 45, Appl
14	211	31.4	46	3	US-09-124-671-1	Sequence 1, Appl1
15	146.5	21.8	889	5	PCT-US93-11725-2	Sequence 2, Appl1
16	146	21.7	35	1	US-08-471-780C-38	Sequence 38, Appl
17	146	21.7	35	1	US-08-467-282B-38	Sequence 38, Appl
18	146	21.7	35	2	US-08-471-282A-38	Sequence 38, Appl
19	146	21.7	35	2	US-08-466-710C-38	Sequence 38, Appl
20	146	21.7	35	3	US-08-468-739C-38	Sequence 38, Appl
21	146	21.7	54	1	US-08-471-780C-44	Sequence 44, Appl
22	146	21.7	54	1	US-08-467-282B-44	Sequence 44, Appl
23	146	21.7	54	2	US-08-471-282A-44	Sequence 44, Appl
24	146	21.7	54	2	US-08-466-710C-44	Sequence 44, Appl
25	146	21.7	54	3	US-08-468-739C-44	Sequence 44, Appl
26	146	21.7	60	1	US-08-471-780C-87	Sequence 87, Appl
27	146	21.7	60	1	US-08-467-282B-87	Sequence 87, Appl
28	146	21.7	60	2	US-08-471-282A-87	Sequence 87, Appl
29	146	21.7	60	2	US-08-466-710C-87	Sequence 87, Appl
30	146	21.7	60	3	US-08-468-739C-87	Sequence 87, Appl
31	145.5	21.7	961	5	PCT-US93-11725-4	Sequence 4, Appl1
32	144	21.4	24	4	US-09-379-297-5	Sequence 5, Appl1
33	121	18.0	46	3	US-09-124-671-5	Sequence 5, Appl1
34	121	18.0	667	2	US-08-718-661-2	Sequence 2, Appl1
35	119	17.0	271	4	US-09-259-991A-17292	Sequence 17292, Appl
36	117	17.4	46	3	US-09-124-671-6	Sequence 6, Appl1
37	116	17.3	261	4	US-09-602-565-34	Sequence 34, Appl
38	113.5	16.9	221	4	US-09-069-023-32	Sequence 22, Appl
39	112	16.7	907	3	US-08-989-299-12	Sequence 12, Appl1
40	111	16.5	679	4	US-09-285-991A-18857	Sequence 18857, Appl
41	110	15.7	2211	4	US-09-738-884-1	Sequence 1, Appl1
42	109	16.2	558	4	US-09-259-991A-17202	Sequence 17202, Appl
43	108	16.1	304	4	US-09-252-991A-23116	Sequence 23116, Appl
44	107	15.9	613	4	US-09-252-991A-22113	Sequence 22113, Appl
45	107	15.9	413	4	US-09-345-473E-39	Sequence 39, Appl1

```

ALIGNMENTS
RESULT 1
US-09-124-671-23
; Sequence 23, Application US/09124671A
; Patent NO. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ. ID NOS.: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
;
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric human COMP-KOBL
US-09-124-671-23

Alignment Scores:
Pred. No.:
2.3e-50 Length: 109
Score: 579.00 Matches: 108
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 86.16% Indels: 0
DB: 3 Gaps: 0

US-09-696-872-24 (1-372) x US-09-124-671-23 (1-109)
16 AGGACACGATTTTACGCTTGCTGCGCCCTTGCGGACAGCTGCGACGCTGCGCAAAAAACA 75
|||||
2 ArgtyrMetIleLeucylcysLeuAlaValCysSerAlaAlaIlySlyGly 230

```



CURRENT APPLICATION NUMBER: US/09/124,671A  
CURRENT FILING DATE: 1998-07-29  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 27  
LENGTH: 109  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chimeric human TSP1-KDEL  
US-09-124-671-27

Alignment Scores:  
Pred. No.: 2,72e-35 Length: 109  
Score: 429.00 Matches: 78  
Percent Similarity: 84.26% Conservative: 13  
Best Local Similarity: 72.22% Mismatches: 17  
Query Match: 63.84% Indels: 0  
DB: 3 Gaps: 0

US-09-696-872-24 (1-372) x US-09-124-671-27 (1-109)

QY 16 AGGTACATGATTTTAAAGCTTCTGCGCCCTTGCGGAGTCTGCAGCGCTGCCAAAAAGCA 75  
Db 2 ArgTyMetIleLeuGlyLeuLeuAlaLeuAlaValCysSerAlaAlaIleGlySerGly 21

QY 76 TCCAGCGCTGGGTGAGACGTGTTCAGACCTGGGCCCCGAGATGCTTCGGGAATCGAG 135  
Db 22 SerSerLeuGlyGlyAspCysGlySerGlyGlnThrValAlaLeuValThrGlnLeuThr 41

QY 136 GAAACCAACCGCGCGCTGCAGACGTGCGGAGCTGCTGCGGACAGTCAGGAGATC 195  
Db 42 LeuHeaenGlnIleLeuValGlnLeuArgAspAlaIleArgAspGlnValIleGlySerGly 61

QY 196 ACGTTCTGAAAAACACGGTATGAGTGTACCGGTCGGGCGCGGACCGCGGAA 255  
Db 62 SerLeuIleArgAsnThrIleMetGlnCysGlnValCysGlyProGlnProGlnProlys 81

QY 256 CCGGACCGCGCGCGCGCGCGCGCGCGGAAACCGGACCGGAAACCGGAAAGTACC 315  
Db 82 ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 101

QY 316 GGATCATCAGAAAAAGATGAGTTG 339  
Db 102 GlySerSerGlnIlyAspGlnLeu 109

RESULT 5  
US-09-124-671-34  
Sequence 34, Application US/09124671A  
Patent No. 6160088  
GENERAL INFORMATION:  
APPLICANT: Rothman, James  
APPLICANT: Mayhew, Mark  
FILE REFERENCE: 31488  
CURRENT APPLICATION NUMBER: US/09/124,671A  
CURRENT FILING DATE: 1998-07-29  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 34  
LENGTH: 134  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: KDEL/myc  
US-09-124-671-34

Alignment Scores:  
Pred. No.: 2.83e-35 Length: 134  
Score: 429.00 Matches: 89  
Percent Similarity: 70.29% Conservative: 8  
Best Local Similarity: 64.49% Mismatches: 9

Query Match: 63.84% Indels: 32  
DB: 3 Gaps: 4

US-09-696-872-24 (1-372) x US-09-124-671-34 (1-134)

QY 10 ATGGGAAGTACATGATTTTA-----GGCTTGCTCGCCCTTGCGGAGTCTGCAGCGCT 63  
Db 1 MetGlyLysPheThrValAlaAlaAlaLeuLeuLeuGlyAlaValArgAlaGln 20

QY 64 GCCAAAAA-----GGA 75  
Db 21 GlySerGlnGlnIlySerGlnIlySerGlnIlySerGlnIlySerGlnIlySerGlnIly 40

QY 76 TCCAGCGCTGGGTGAGACGTGTTCAGACCTGGGCCCCGAGATGCTTCGGGAATCGAG 135  
Db 41 SerSerLeuGlyGlyAspCys-----ProGlnMetLeuArgGlnIlyGln 56

QY 136 GAAACCAACCGCGCGCTGCAGACGTGCGGAGCTGCTGCGGACAGTCAGGAGATC 195  
Db 57 GlnThrAsnAlaAlaLeuGlnAspValArgGlnLeuLeuArgGlnIlyGlnIly 76

QY 196 ACGTTCTGAAAAACACGGTATGAGTGTACCGCTCGCGG----- 237  
Db 77 ThrPheLeuIlyAsnThrValMetGlnCysAspAlaCysGlyMetGlnProAlaArgThr 96

QY 238 -----CCGACCGCGCGCGGAAACCGGACCGGACCGGACCGGACCGGAA 285  
Db 97 ProGlyThrSerProGlnProGlnProGlnProGlnProGlnProGlnProGlnProlys 116

QY 286 CCGGACCGGAAACCGGAAACCGGAAAGTACCGGATCATCAGAAAAAGATGAGTTG 339  
Db 117 ProGlnProlysProGlnProGlnIlyThrGlySerSerGlnIlyAspGlnLeu 134

RESULT 6  
US-09-124-671-13  
Sequence 13, Application US/09124671A  
Patent No. 6160088  
GENERAL INFORMATION:  
APPLICANT: Rothman, James  
APPLICANT: Mayhew, Mark  
FILE REFERENCE: 31488  
CURRENT APPLICATION NUMBER: US/09/124,671A  
CURRENT FILING DATE: 1998-07-29  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 13  
LENGTH: 115  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chimeric rat comp  
US-09-124-671-13

Alignment Scores:  
Pred. No.: 1.24e-34 Length: 115  
Score: 422.50 Matches: 88  
Percent Similarity: 77.87% Conservative: 7  
Best Local Similarity: 72.13% Mismatches: 8  
Query Match: 62.87% Indels: 19  
DB: 3 Gaps: 4

US-09-696-872-24 (1-372) x US-09-124-671-13 (1-115)

QY 10 ATGGGAAGTACATGATTTTA-----GGCTTGCTCGCCCTTGCGGAGTCTGCAGCGCT 63  
Db 1 MetGlyLysPheThrValAlaAlaAlaLeuLeuLeuGlyAlaVal----- 17

QY 64 GCCAAAAAGGATCAGCGCTGGGTGAGACTGTTTTCAGACCTGGGCCCCGAGATGCTT 123  
Db 18 ArgAlaGlnIlySerSerLeuGlyIly-----AspLeuAlaProGlnMetLeu 33

```

OY      124  CGGGAACATCGCAGGAAACCAACCGCGCGCTGCGACGACGTCCGGACATCTGCTGCGCGACGAG 183
Db      34   ArgGtIuLeuGIngluThrAsnAlaAlaLeuGInAspValArgGluLeuLeuArgGInGIn 53
OY      184  GTCAAGGAGATCACTTCTGTGAAAAACAAGGATAGTGAAGTGTGACCGCTGCGCGG----- 237
Db      54   VallysGluIleThrPheLeuYsAsnThrValMetGluCysAspAlaCysGlyMetGln 73
OY      238  -----CCGACGCCGCGACCCGAAACCGCAGCCGCGACGCGCGAG 273
Db      74   ProAlaArgThrProGlyThrSerProGlnProGlnProLysProGlnProGlnProGln 93
OY      274  CGCAGCCGGAACCGCAGCCGGAACCGGAAACCGGAAGTAAACCGGATCATCAGAAAAAGAT 333
Db      94   ProGlnProLysProGlnProLysProGlnProGlnProGlnGlyThrIleYsSerGlnLysAsp 113
OY      334  GAGTTG 339
Db      114  GluLeu 115

RESULT 7
US-09-124-671-21
; Sequence 21, Application US/09124671A
; Patent No. 6160088
; GENERAL INFORMATION:
; APPLICANT: Rothman, James
; APPLICANT: Mayhew, Mark
; APPLICANT: Hoe, Mee
; TITLE OF INVENTION: KDEL RECEPTOR INHIBITORS
; FILE REFERENCE: 31488
; CURRENT APPLICATION NUMBER: US/09/124,671A
; CURRENT FILING DATE: 1998-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric Xenopus laevis TSP4-KDEL
US-09-124-671-21

Alignment Scores:
Pred. No.: 2.57e-30 Length: 109
Score: 379.50 Matches: 70
Percent Similarity: 77.68% Conservative: 17
Best Local Similarity: 62.50% Mismatches: 20
Query Match: 56.47% Indels: 5
DB: 3 Gaps: 2

US-09-696-872-24 (1-372) x US-09-124-671-21 (1-109)
OY      10  ATGGGAGATCATGATATTTTA-----GGCTTGCTCGCCCTTGCGGCGAGTCTGCGACGCT 63
Db      1  MetGlyLysPheThrValValAlaAlaAlaLeuLeuLeuGlyAlaVal----- 17
OY      64  GCCAAAAGAAGATCCAGCGCTGGGTGGAGACTGTGTTCTACAGCTGGGCGCGGAGATGCTT 123
Db      18  ArgAlaGluGlySerSerLeuGlnIleGlyAspCysGlyAlaSerValSerArgGlnLeuIle 37
OY      124  CGGGAATCGCAGGAAACCAACCGCGCGCTGCGACAGCTGCGGAGATCTGCGTGGCGGACGAG 183
Db      38  GlyIleIleThrGlnMetLeuAsnGlnMetLeuGlyGluLeuArgAspValMetArgGlnGln 57
OY      184  GTCAGGAGATCACTTCTGTGAAAAACAAGGATAGTGAAGTGTGACCGCTGCGCGGCGGAG 243
Db      58  VallysGluThrIleMetPheLeuArgAsnThrIleAlaGluCysGlnAlaCysGlyProGln 77
OY      244  CCGCAGCCGGAACCGCAGCCGCGACGCGCAGCCGGAACCGGAAACCGGACCGGAAACCGGAA 303
Db      78  ProGlnProLysProGlnProGlnProGlnProGlnProGlnProLysProGlnProLysProGln 97
OY      304  CCGGAAGGTACCGGATCATCAGAAAAAGATGAGTTG 339

```

[illegible]







COMPUTER: IBM-compatible  
OPERATING SYSTEM: MS-DOS Version 3.3  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/11725  
FILING DATE: filed herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/985,296  
FILING DATE: 04-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, Edward R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: B0801/7005WO  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 889 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: yes  
PCT-US93-11725-2

Alignment Scores:  
Pred. No.: 9.79e-07 Length: 889  
Score: 146.50 Matches: 44  
Percent Similarity: 47.33% Conservative: 18  
Best Local Similarity: 33.59% Mismatches: 36  
Query Match: 21.80% Indels: 33  
DB: 5 Gaps: 5

US-09-696-872-24 (1-372) x PCT-US93-11725-2 (1-889)

```
OY 48 GGCAGTCTGCAGCGCTGCCAATAAAAGATCCAGCTGGGTGAGACTGTGTTCAGACCT 107
Db 132 G1YAla11eG1ng1uCysPheMeCglN-LysSerGlu1aG1yG1nG1nThrG1yAepVa 151
OY 108 GGGCCCGAGATGCTTCGGGAAGTGCAGAAACCAACCGCGCTGCAGAGACGTGCGGA 167
Db 151 lSerArg1nleu1leG1yG1n1leThrG1nMeCAnG1nMe1leuG1yG1nleuArgAs 171
OY 168 CTGCTCGCGCAGCAGGTGCAGGAGATCAGCTTCTGAAAAACACCGGTGATGAGTGA 227
Db 171 pValMeCArg1nG1nVal1ySG1uThrMeCPhelEuArgAnThr1leAlaG1uCysG1 191
OY 228 CGCGTGGGGCGCGCAGCGCAG---CCGAACCGCAGCGCGCAG----- 273
Db 191 nAlaCysG1yLeuG1yProAspPheProLeuProThr1yValProG1nArgLeuAlaTh 211
OY 274 ---CCGCAGCCGAACCG----- 288
Db 211 rThrThrProPro1ySProArgCysAspAlaThrSerCysPheArgG1yValArgCys11 231
OY 289 -----CAGCCGAACCGGAACCGGAAGT---ACCGATCA----- 321
Db 231 eaSPThrG1uG1yG1PheG1nCysG1yProCysProG1uG1yTyTrThrG1yAanG1yVa 251
OY 322 -----TCAGAAAAAGATGAGTGTAGG 343
Db 251 11leCysThraPvalaepGlu-CysArg 260
```

Search completed: August 26, 2003, 16:26:26  
Job time : 17.5 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_n2p model

Run on: August 26, 2003, 16:24:54 ; Search time 41.5 Seconds

(without alignments)  
2365.772 Million cell updates/sec

Title: US-09-696-872-24

Sequence: 1 aagcttaccatggaagta.....aattcatatgcattccgag 372

Scoring table:

BIOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Rgapop 6.0 , Rgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 497079 seqs, 131961718 residues

Total number of hits satisfying chosen parameters: 994158

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_n2p,model -DEV=xdp  
-O=/cg21/USPTO.spool/US09696872/runat\_26082003\_150545\_8814/app\_query.fasta.1.519  
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=bl0sum62  
-TRANS=human40.cdd -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MALLEN=2000000000 -USER=US09696872@cgn\_1.1.83@runat\_26082003\_150545\_8814  
-NCEPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

1	262.5	39.1	242	10	US-09-919-603-5	Sequence 5, Appl1
2	262.5	39.1	300	10	US-09-919-603-7	Sequence 7, Appl1
3	262.5	39.1	757	12	US-09-919-603-3	Sequence 3, Appl1
4	262.5	39.1	757	12	US-10-301-822-41	Sequence 41, Appl1
5	262.5	39.1	757	12	US-10-177-293-72	Sequence 72, Appl1
6	145.5	21.7	961	15	US-10-017-721-4	Sequence 4, Appl1
7	140	20.8	178	15	US-10-154-971-22	Sequence 22, Appl1
8	139	20.7	303	15	US-10-154-971-24	Sequence 24, Appl1
9	126.5	18.8	578	15	US-10-156-761-12544	Sequence 12544, A
10	122	18.2	400	12	US-10-301-822-155	Sequence 155, App
11	118.5	16.9	2456	10	US-10-184-644-57	Sequence 57, Appl1
12	118.5	16.9	2456	15	US-10-184-634-57	Sequence 57, Appl1
13	117	17.4	285	10	US-09-881-752A-228	Sequence 228, Appl1
14	116	17.3	261	15	US-10-280-953-17	Sequence 17, Appl1
15	115.5	16.5	1547	12	US-10-140-472-417	Sequence 417, App
16	115.5	16.5	1547	12	US-10-123-155-417	Sequence 417, App
17	115.5	16.5	1547	12	US-10-123-155-417	Sequence 417, App
18	115.5	16.5	1547	16	US-10-146-731-417	Sequence 417, App
19	115	16.4	2498	12	US-10-140-472-483	Sequence 483, App
20	115	16.4	2498	12	US-10-141-761-483	Sequence 483, App
21	115	16.4	2498	15	US-10-123-155-483	Sequence 483, App
22	115	16.4	2498	16	US-10-146-731-483	Sequence 483, App
23	113.5	16.2	2868	12	US-10-140-472-119	Sequence 119, App
24	113.5	16.2	2868	12	US-10-141-761-119	Sequence 119, App
25	113.5	16.2	2868	15	US-10-123-155-119	Sequence 119, App
26	113.5	16.2	2868	16	US-10-146-731-119	Sequence 119, App
27	113	16.1	1773	15	US-10-184-634-555	Sequence 555, App
28	113	16.1	1773	15	US-10-184-634-555	Sequence 555, App
29	112.5	16.0	3554	12	US-10-140-472-537	Sequence 537, App
30	112.5	16.0	3554	12	US-10-141-761-537	Sequence 537, App
31	112.5	16.0	3554	15	US-10-123-155-537	Sequence 537, App
32	112.5	16.0	3554	16	US-10-146-731-537	Sequence 537, App
33	111.5	15.6	807	14	US-10-138-221-7	Sequence 7, Appl1
34	111.5	15.9	3305	15	US-10-184-644-87	Sequence 87, Appl1
35	111.5	15.9	3305	15	US-10-184-634-87	Sequence 87, Appl1
36	111	15.8	2403	15	US-10-184-644-45	Sequence 45, Appl1
37	111	15.8	2403	15	US-10-184-644-45	Sequence 45, Appl1
38	110.5	15.7	2849	12	US-10-140-472-285	Sequence 285, App
39	110.5	15.7	2849	12	US-10-141-761-285	Sequence 285, App
40	110.5	15.7	2849	15	US-10-123-155-285	Sequence 285, App
41	110.5	15.7	2849	16	US-10-146-731-285	Sequence 285, App
42	110	15.7	2211	14	US-10-096-961-1	Sequence 1, Appl1
43	109.5	15.6	1985	12	US-10-140-472-143	Sequence 143, App
44	109.5	15.6	1985	12	US-10-141-761-143	Sequence 143, App
45	109.5	15.6	1985	15	US-10-123-155-143	Sequence 143, App

#### ALIGNMENTS

Database : Published Applications AA.\*

- 1: /cg2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cg2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cg2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cg2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cg2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cg2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cg2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cg2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cg2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cg2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 11: /cg2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cg2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cg2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cg2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cg2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cg2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 17: /cg2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cg2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query	Score	Match length	ID	Description
------------	-------	-------	--------------	----	-------------

RESULT 1

US-09-919-603-5

Sequence 5, Application US/09919603

Patent No. US20020137679A1

GENERAL INFORMATION:

APPLICANT: Lawler, John W.

TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP

FILE REFERENCE: 1440.1035-007

CURRENT FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: PCT/US00/02482

PRIOR FILING DATE: 2000-02-01

PRIOR APPLICATION NUMBER: 60/118,053

PRIOR FILING DATE: 1999-02-01

NUMBER OF SEQ ID NOS: 21

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5

LENGTH: 242

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: chimeric protein

US-09-919-603-5

**Alignment Scores:**

Pred. No.:	7.86e-16	length:	24
Score:	262.50	Matches:	57
Percent Similarity:	75.61%	Conservative:	5
Best Local Similarity:	69.51%	Mismatches:	9
Query Match:	39.06%	Indels:	11
DB:	10	Gaps:	2

US-09-696-872-24 (1-372) X US-09-919-603-5 (1-242)

Qy	34	TTGCTTCGCCCTTGGCGCAGCTCTGCAGCCGTCACAAAAAGATCCAGCTCGGTGGAGAC	93
Db	10	LeuLeuThrLeuAlaAlaLeuGlyAlaSerGlyAlaGlnGlyGlnSerProLeuGly-----	27
Qy	94	TGTTGTTCAGACTGGGGCCCGCAGATGCTTGGGAACTGCAGAGAAACCAACGCGCGCTG	15
Db	28	-----SerAspLeuGlyProGlnMetLeuArgIuLeuGlnGluThrAsnAlaAlaLeu	45
Qy	154	CAGACGTGCGGGACTGAGCTGGCGCAGCAGGTCAAGGAGATCAGCTTCTGAAAAACG	213
Db	46	GlnAspValAlaArgAspTrpLeuArgGlnGlnValAlaArgIuLLeuPheLeuValAsnThr	65
Qy	214	GTGATGAGTGTGACCGCGTGGCGCGCCAG-----CCGACGCGG	252
Db	66	ValMetGluCysAspAlaCysGlyMetGlnGlnSerValAlaArgThrGlyLeuProSerVal	85
Qy	253	AAACCG	258
Db	86	ArgPro	87

RESULT 2  
US-09-91

Sequence 7, Application US/09919603  
Patent No. US20020137679A1

```

? GENERAL INFORMATION:
? APPLICANT: Lawler, John W.
? TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSH
? TITLE OF INVENTION: Chimeric Proteins
? FILE REFERENCE: 1440.1033-007
? CURRENT APPLICATION NUMBER: US/09/919, 603
? CURRENT FILING DATE: 2001-07-30
? PRIOR APPLICATION NUMBER: PCT/US00/02482
? PRIOR FILING DATE: 2000-02-01
? PRIOR APPLICATION NUMBER: 60/118, 053
? PRIOR FILING DATE: 1999-02-01
? NUMBER OF SEQ. ID NOS: 21
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 7
? LENGTH: 300
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: chimeric protein
? IS-09-919-603-7

```

**Alignment Scores:**

Pred. No.:	8.04e-16	Length:	300
Score:	262.50	Matches:	57
Percent Similarity:	75.61%	Conservative:	5
Best Local Similarity:	69.51%	Mismatches:	9
Query Match:	39.06%	Indels:	1
DB:	10	Gaps:	2

US-09-696-872-24 (1-372) X US-09-919-603-7 (1-300)

QY 3 TTGCTGACCCCTTGGCGCATCTCTGACGGCTGCAAAAAGATCCAGCTGGGTGGAGAC 93  
 Db 10 LeuLeuThrLeuAlaAlaLeuGlyAlaSerGlyGlnGlyGlnSerProLeuGly----- 27  
 QY 94 TGTATTTCAGACTCTGGGCCCCGACAGATGCTTTGGGAACTGCAGAGAAACCAATCGCGGCTG 155  
 Db 28 -----SerAspIleuGlyProGlnInntLeuAlaGlyIleuGlnGlnGlyIleuThrAsnAlaAlaLeu 45

QY 154 CAGGACGTGCGGACTGGCTGCGGCAGCAGGTCAAGGAGATCACGTTCTGAAACACAG 213

Db 46 GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuLysAsnThr 65

214 GTGATCGAGTGTGACGCGTGGCGGCGCAG-----CCGCAAGCCG 252

Db 66 ValMetGluCysAspAlaCysGlyMetGlnGlnSerValArgThrGlyLeuProSerVal 85

QY 253 AAACCG 258  
: : : |||

Db 86 ArgPro 87

RESULT 3  
US-09-91

; Sequence 3, Application US/09919603  
; Patent No. US20020137679A1

```

? GENERAL INFORMATION:
? APPLICANT: Lawyer, John W.
? TITLE OF INVENTION: COMP/TSP-1, COMP/TSP-2 and Other TSP
? TITLE OF INVENTION: Chimeric Proteins
? FILE REFERENCE: 1440.1033-007
? CURRENT APPLICATION NUMBER: US/09/919,603
? CURRENT FILING DATE: 2001-07-30
? PRIOR APPLICATION NUMBER: PCT/US00/02482
? PRIOR FILING DATE: 2000-02-01
? PRIOR APPLICATION NUMBER: 60/118,053
? PRIOR FILING DATE: 1999-02-01
? NUMBER OF SEQ. ID NOS: 21
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ. ID NO 3
? LENGTH: 757
? TYPE: PRT
? ORGANISM: Homo sapiens
? IS-09-919-603-3

```

**Alignment Scores:**

Pred. No.:	8.85e-16	Length:	75
Score:	263.50	Matches:	57
Percent Similarity:	75.61%	Conservative:	5
Best Local Similarity:	69.51%	Mismatches:	9
Query Match:	39.06%	Indels:	11
DB:	10	Gaps:	2

US-09-696-872-24 (1-372) x US-09-919-603-3 (1-757)

[illegible]

RESULT 4

US-10-301-822-41  
; Sequence 41, Application US/10301822  
; Publication No. US20030148410A1  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc  
; APPLICANT: Berger, Allison

```

1  APPLICANT: Guillemette, Tracy L.
2  APPLICANT: Kamatkar, Shubhangi
3  APPLICANT: Schlegel, Robert
4  APPLICANT: Monahan, John E.
5  APPLICANT: Tribodeau, Stephen N.
6  APPLICANT: Burgart, Lawrence J.
7  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
8  TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
9  TITLE OF INVENTION: THERAPY OF COLON CANCER
10 FILE REFERENCE: MPM01-029P2RNM
11 CURRENT APPLICATION NUMBER: US/10/301,822
12 CURRENT FILING DATE: 2002-11-21
13 PRIOR APPLICATION NUMBER: US 60/339,971
14 PRIOR FILING DATE: 2001-12-10
15 PRIOR APPLICATION NUMBER: US 60/361,978
16 PRIOR FILING DATE: 2002-03-05
17 PRIOR APPLICATION NUMBER: US 60/381,988
18 PRIOR FILING DATE: 2002-05-20
19 NUMBER OF SEQ. ID NOS: 228
20 SOFTWARE: FastSeq for Windows Version 4.0
21 SEQ ID NO 41
22 LENGTH: 757
23 TYPE: PR1
24 ORGANISM: Homo Sapiens
25 US-10-301-822-41

```

Alignment Scores:	
Pred. No.:	8,85e-16
Score:	263.50
Percent Similarity:	75.61%
Best Local Similarity:	69.51%
Query Match:	39.06%
BB:	12
Length:	75
Matches:	57
Conservative:	5
Mismatches:	1
Indels:	11
Gaps:	1

US-09-696-872-24 (1-372) X US-10-301-822-41 (1-757)

QY	3	TTGCTCGCCCTTGCGGCAGCTCTGCAGAAAAGATCCAGCTGGGTGGAGAC	93
		::::   ::  ::::	
Db	10	LeuLeuThrLeuAlaAlaLeuGlyAlaSerGlyGlnSerProLeuGlyI-	27
QY	94	TGTTGTTCAGACCTGGGCCCGCAGATGCTTCGGGAATGCAGAAACCAACGCGCGCTG	153
Db	28	-SerAspLeuGlyProGlnMetLeuArgIuLeuGlnGluThrAsnAlaAlaLeu	45
QY	154	CAGGAGTGTCCGGGACTGTGGCGCAGCAGGTCCAGGGAGATCACTTCTTAAAAACAG	213
Db	46	GlnAspValArgAspTrpLeuArgGlnGlnValArgGluIleThrPheLeuIysAsnThr	65
QY	214	GTGATAGAGTGTGACCGCTGCGGGCGCGCAG-----CCGACACCG	252
Db	66	ValMetGluCysAspAlaCysGlnMetGlnGlnSerValArgThrGlyLeuProSerVal	85
QY	253	AAACCG	258
Db	86	ArgPro	87

RESULT 5  
US-10-17

Sequence 72, Application US/10177293  
Publication No. US20030124128A1

APPLICANT: Lillie, James  
APPLICANT: Glatt, Karen  
APPLICANT: Zho, Xunlei  
APPLICANT: Ganeshvarpu, Manjula  
APPLICANT: Kamathkar, Shubhang  
APPLICANT: Mettens, Maureen  
APPLICANT: Meyer, Vic  
APPLICANT: Wang, Youzhen  
APPLICANT: Xu, Yongyao  
APPLICANT: Hoersch, Sebastian  
APPLICANT: Monahan, John  
APPLICANT: Meyers, Rachel E.

```

APPLICANT: Bast Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Puzstai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 72
LENGTH: 757
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-72

```

Alignment Scores:	
Pred. No.:	8, 85e-16
Score:	263.50
Percent Similarity:	75.61%
Best Local Similarity:	69.51%
Query Match:	39.06%
DB:	15
Length:	75
Matches:	57
Conservative:	5
Mismatches:	9
Indels:	11
Gaps:	

US-09-696-872-24 (1-372) X US-10-177-293-72 (1-757)

QY 3 TTTGCTCCGCCCTTTGGGGAGATCTCGACGGCCTGCACAAAAAGGATCTCACCTCTGGGTGGAGAC 93  
 Db 10 LeuLeuThrLeuAlaIaIaLeuLeuYAlaSerGlyGlnGlyGlnSerProLeuGlyI----- 27  
 QY 94 TGTGTTTCAGACTCTGGGCCCGGCAGATGCTTCGGGAACATGCAGAAACCAACGCCGCGCTG 153  
 Db 28 -----SerAspLeuGlyProGlnMetLeuArgIuLeuGlnGlnIuThrAsnAlaIaLeu 45  
 QY 154 CAGGACGTGCGCGGAACTTGCTGCTGCGCAGAGAGTCAAGGAGATCAGACGTTCCGAAAAACAG 213  
 Db 46 GlnAspValaArgAspTrpLeuArgGlnGlnIuValaArgIuIleThrPheLeuLYAsnThr 65  
 QY 214 GTGATGGAATGTGAACGCGCTGCGCGCCGAG-----CCGACACCG 252  
 Db 66 ValMetGluCyAspAlaCysGlnMetClnGlnSerValaArgThrGlyLeuProSerVal 85  
 QY 253 AAACCG 258  
 Db 86 ArgPro 87

## RESULT 6

US-10-01/-721-4  
; Sequence 4, Application US/10017721  
; Publication No. US20030096248A1

GENERAL INFORMATION: Jeanette  
APPLICANT: McCarthy, George  
APPLICANT: Daley, George  
APPLICANT: Bolk, Stacy  
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF VASCULAR DISEASE  
FILE REFERENCE: MMI-003  
CURRENT APPLICATION NUMBER: US/10/017,721  
CURRENT FILING DATE: 2001-12-14

PRIOR APPLICATION NUMBER: US 60/317,033  
PRIOR FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: US 60/330,248  
PRIOR FILING DATE: 2001-10-17  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 961  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-017-721-4

Alignment Scores:  
Pred. No.: 4,376-05 Length: 961  
Score: 145.50 Matches: 30  
Percent Similarity: 57.14% Conservative: 10  
Best Local Similarity: 42.86% Mismatches: 25  
Query Match: 21.65% Indels: 5  
DB: 15 Gaps: 2

US-09-696-872-24 (1-372) x US-10-017-721-4 (1-961)

QY 103 GACCTGGGCGCCGACATGCTTCGGAGACTGCAGAAACCGCGCGCTGCAGACGCG 162  
DB 218 AaPheaaAaAaGlnPheleuGlyGlnMetThrGlnLeuAaGlnLeuLeuGlyGlnVal 237  
QY 163 CGGAGCTGCTGCGGAGCAGCAGATCAGGATCAGTTCTGAAAAACAGCGTGTAGAG 222  
DB 228 LysaAaPheleuAaAaGlnGlnValLysGlnThrSerPheleuAaAaThrLleAaGlu 257  
QY 223 TGTGACGCTGCGGCG 270  
DB 258 CysGlnAaCyGlyPheleuLysPheGlnSerProThrPheValValaPro 277  
QY 271 CAGCGCAGCGGAAACCGCAGCGCGGAAACCG 300  
DB 278 AlaPro---ProAlaProProThrArgPro 286

RESULT 7  
US-10-154-971-22  
Sequence 22, Application US/10154971  
Publication No. US20030088074A1  
GENERAL INFORMATION:  
APPLICANT: Hamers, Raymond  
TITLE OF INVENTION: VARIABLE FRAGMENTS OF IMMUNOGLOBULINS -  
USE FOR THERAPEUTIC OR VETERINARY PURPOSES  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESSES:  
ADDRESS: SPENCER & FRANK  
STREET: 1100 New York Avenue, N.W., Suite 300 East  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/154,971  
FILING DATE: 28-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/945,244  
FILING DATE: <Unknown>  
APPLICATION NUMBER: PCT/EP96/01725  
FILING DATE: 25-APR-1996  
APPLICATION NUMBER: EP 95400932.0  
FILING DATE: 25-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Gollin, Michael A.

REGISTRATION NUMBER: 31,957  
REFERENCE/DOCKET NUMBER: GUPLA 0003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-414-4000  
TELEFAX: 202-414-4040  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 178 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 22:  
US-10-154-971-22

Alignment Scores:  
Pred. No.: 0,000117 Length: 178  
Score: 140.00 Matches: 24  
Percent Similarity: 88.89% Conservative: 0  
Best Local Similarity: 88.89% Mismatches: 3  
Query Match: 20.83% Indels: 0  
DB: 15 Gaps: 0

US-09-696-872-24 (1-372) x US-10-154-971-22 (1-178)

QY 228 CCGCAGCCGAGCGGAAACCGCAGCGCGGAGCGCGGAGCGCGGAAACCGCAGCGGAA 297  
DB 138 ProGlnProGlnProLysProGlnProGlnProGlnProGlnProLysProGlnProLys 157  
QY 298 CCGGAAACCGGAAAGGTACCGGA 318  
DB 158 ProGlnProMetGluCyGly 164

RESULT 8  
US-10-154-971-24  
Sequence 24, Application US/10154971  
Publication No. US20030088074A1  
GENERAL INFORMATION:  
APPLICANT: Hamers, Raymond  
TITLE OF INVENTION: VARIABLE FRAGMENTS OF IMMUNOGLOBULINS -  
USE FOR THERAPEUTIC OR VETERINARY PURPOSES  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESSES:  
ADDRESS: SPENCER & FRANK  
STREET: 1100 New York Avenue, N.W., Suite 300 East  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/154,971  
FILING DATE: 28-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/945,244  
FILING DATE: <Unknown>  
APPLICATION NUMBER: PCT/EP96/01725  
FILING DATE: 25-APR-1996  
APPLICATION NUMBER: EP 95400932.0  
FILING DATE: 25-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Gollin, Michael A.  
REGISTRATION NUMBER: 31,957  
REFERENCE/DOCKET NUMBER: GUPLA 0003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-414-4000  
TELEFAX: 202-414-4040  
INFORMATION FOR SEQ ID NO: 24:









**This Page Blank (usp10)**

[illegible]

15	118	17.6	422	2	E91067	hydrophobic prote
16	117	17.4	285	2	B64687	siderophore-mediat
17	117	17.4	309	2	T17557	procyclin homolog
18	117	17.4	497	1	F83634	hypothetical prote
19	116.5	17.3	115	1	FAUTPC	procyelic acidic r
20	116.5	17.3	143	2	A26938	procyclin acidica
21	116.5	17.3	145	2	A44418	procyclin PSSA-1 -
22	115	17.1	280	2	F71829	siderophore-mediat
23	114	17.0	266	2	T44781	conb protein limpo
24	113.5	16.9	127	2	S09372	hypothetical prote
25	113.5	16.9	139	2	S70010	hypothetical prote
26	113.5	16.9	221	2	S70009	glutamate/proline-
27	113	16.8	407	2	B82478	hypothetical prote
28	112.5	16.7	839	1	TQZMCA	probable transposa
29	112.5	16.7	949	2	T06858	hypothetical prote
30	112.5	16.7	1528	2	D85912	hypothetical prote
31	112.5	16.7	1569	2	A65044	hypothetical prote
32	112.5	16.7	1571	2	C91068	hypothetical prote
33	112	16.7	141	2	T09546	extensin like prot
34	112	16.7	203	2	T16348	hypothetical prote
35	112	16.7	907	2	T15792	hypothetical prote
36	111.5	16.6	138	2	B27863	Ac hypothetical pr
37	111.5	16.6	807	2	T02916	hypothetical prote
38	111	16.5	427	2	T03935	probable transposa
39	111	16.5	1749	2	S75138	hypothetical prote
40	110.5	16.4	149	2	B96651	protein T3P18.7 (l
41	110.5	16.4	179	2	T34145	hypothetical prote
42	110	16.4	792	2	F90566	conserved hypothet
43	109.5	16.3	221	2	B43701	PR24 protein - ch
44	109	16.2	197	2	T18918	hypothetical prote
45	109	16.2	312	2	T28085	hypothetical prote

## ALIGNMENTS

## RESULT 1

cartilage oligomeric matrix protein precursor - rat  
 N:Alternate names: thrombospondin homolog COMP  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 30-Apr-1993 #sequence\_revision 03-May-1994 #text\_change 02-Aug-2002  
 C/Accession: A44315  
 R:Oldberg, A.; Antonsson, P.; Lindblom, K.; Heinegard, D.  
 J. Biol. Chem. 267, 22346-22350, 1992  
 A>Title: COMP (cartilage oligomeric matrix protein) is structurally related to the thromb  
 A:Reference number: A44315; MUID:93054522; PMID:1429587  
 A:Accession: A44315  
 A:Molecule type: mRNA  
 A:Residues: 1-755 <OLD>  
 A:Cross-references: GB:X79214; NID:9297438; PIDD:CAA51419.1; PID:9297439  
 A:Experimental source: tracheal chondrocyte  
 A>Note: sequence extracted from NCBI backbone (NCBI:P.11022)  
 C:Superfamily: thrombospondin 3; EGF homology  
 C:Keywords: pentamer  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:182-219/Domain: EGF homology <EGF>

**Alignment Scores:**

Pred. No.:	6,48e-13	Length:	75
Score:	234.00	Matches:	51
Percent Similarity:	72.84%	Conservative:	8
Best Local Similarity:	62.96%	Mismatches:	18
Query Match:	34.82%	Indels:	4
DB:	2	Gaps:	1

US-09-696-872-24 (1-372) X A44315 (1-755)

34 TTGCTCGCCCTTGGCGAGTCTGCAGCGCTGCCAAAAAGGATCCAGCCTGGGTGAGAC 93

Db 9 valleua laleua l a l a l e u a r g l a t h r g l y g l n i l e p r o l e u g l y c y l y --- 27

94 TGTTCAGACCTGGCCCGCAGATGCTTCGGAACTGCAGGAACCAACGGCGCTG 153

```

Db -----AspleuAlaProGlnMetLeuArgGluLeuGlnGluThrAsnAlaLeu 44
QY 154 CAGCAGTCGCGGAGCTGCTGCGGCGACGATCAGGAGATCATGTTCTGAAAAACAGC 213
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 45 GlnAapValArgGluLeuLeuArgHisArgValValGluIleThrPheLeuYsaThr 64
QY 214 GTGATGAGATGTGACGCGCTGCGGCGCGACCGCAACCGCAGCCGACCGCAG 273
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 65 ValMetGluCysAapAlaCysGlyMetGlnProAlaArgThrProGlyLeuSerValArg 84
QY 274 CCG 276
    |||
Db 85 Pro 85

RESULT 2
A45441
Thrombospondin 4 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jan-2000
C:Accession: A45441
R:Lawler, J.; Duquette, M.; Whitaker, C.A.; Adams, J.C.; McHenry, K.; Desimone, D.W.
J. Cell Biol. 120, 1059-1067, 1993
A:Title: Identification and characterization of thrombospondin-4, a new member of the th
A:Reference number: A45441; MUID:93163109; PMID:8432726
A:Accession: A45441
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-955 <LAN>
A:Cross-references: GB:Z19091; NID:9288777; PIDN:CAAT9518.1; PID:9288778
A>Note: sequence extracted from NCBI backbone (NCBIN:124858, NCBIF:124860)
F:285-319/Domain: EGF homology <EGF>

Alignment Scores:
Pred. No.: 3 3e-05 Length: 955
Score: 146.50 Matches: 44
Percent Similarity: 47.33% Conservative: 18
Best Local Similarity: 33.59% Mismatches: 36
Query Match: 21.80% Indels: 33
DB: 2 Gaps: 5

US-09-696-872-24 (1-372) x A45441 (1-955)
QY 48 GGCAGCTGTGACGCGCTGCCAAAAAGATCCAGCTGTGAGAGATGTTGTTACAGCT 107
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 198 GlyAlaIleGlnGlnGlnCysPheMetGln-LysSerGluAlaGlyGlnGlnThrGlyAspVa 217
    |||||:::|||||:::|||||:::|||||:::|||||
QY 108 GGGCCCGCAGATGCTTCGGGAACTGCAGGAAACCAACGCGGCTGCGAGGAGTGGCGGA 167
    |||||:::|||||:::|||||:::|||||:::|||||
Db 217 IserArgGlnLeuIleGlyGlnIleThrGlnMetAsnGlnMetLeuGlyGluLeuArgAs 237
    |||||:::|||||:::|||||:::|||||:::|||||
QY 168 CTGGCTGCGGCGACGAGTCCAGGAGATCATCGTTCCTGAAAAACAGCGATGAGGTGA 227
    |||||:::|||||:::|||||:::|||||:::|||||
Db 237 PValMetArgGlnGlnValValGlnThrMetPheLeuArgAsnThrIleAlaGluCysGcl 257
    |||||:::|||||:::|||||:::|||||:::|||||
QY 228 CGCGTGGCGGCGCGACCGCAG---CCGAAACCGCAGCCGCGCAG--- 273
    |||||:::|||||:::|||||:::|||||:::|||||
Db 257 nAlaCysGlyLeuGlyProAspPheProLeuProThrLysValProGlnArgLeuAlaThr 277
    |||||:::|||||:::|||||:::|||||:::|||||
QY 274 ---CCGACCGCGAAACCG--- 288
    |||||:::|||||:::|||||:::|||||:::|||||
Db 277 rThrThrProProLysProArgCysAapAlaThrSerCysPheArgGlyValArgCys11 297
    |||||:::|||||:::|||||:::|||||:::|||||
QY 289 ---CACCGAAACCGGACCGGAAAGT---ACCGGATC----- 321
    |||||:::|||||:::|||||:::|||||:::|||||
Db 297 eaSpThrGluGlyGlyPheGlnCysGlyProCysProGluGlyTyThrGlyAsnGlyVa 317
    |||||:::|||||:::|||||:::|||||:::|||||
QY 322 ---TCAGAAAAAGATGACTGTAGG 343
    |||||:::|||||:::|||||:::|||||:::|||||
Db 317 lIleCysThrAspValAspGlu-CysArg 326
    |||||:::|||||:::|||||:::|||||:::|||||

RESULT 3
S33598*

```

```

Ig gamma-2 chain - Arabian camel (fragment)
C:Species: Camelus dromedarius (Arabian camel)
C:Date: 19-Mar-1997 #sequence_revision 10-Oct-1997 #text_change 11-Jan-2000
C:Accession: S33598
R:Hamers-Casterman, C.; Alarouch, T.; Muyldermans, S.; Robinson, G.; Hamers, C.; Songa,
Nature 363, 446-448, 1993
A:Title: Naturally occurring antibodies devoid of light chains.
A:Reference number: S33598; MUID:93275410; PMID:8502296
A:Accession: S33598
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-135 <HAM>
A:Experimental source: spleen
C:Superfamily: immunoglobulin V region; immunoglobulin homology
C:Keywords: disulfide bond; immunoglobulin
F:14-62/Disulfide bonds: #status predicted

Alignment Scores:
Pred. No.: 5 07e-05 Length: 135
Score: 146.00 Matches: 25
Percent Similarity: 96.15% Conservative: 0
Best Local Similarity: 96.15% Mismatches: 1
Query Match: 21.73% Indels: 0
DB: 2 Gaps: 0

US-09-696-872-24 (1-372) x S33598 (1-135)
QY 238 CCGGACCGCGACCGGAAACCGGACCGCGACCGCGACCGGAAACCGGACCGGAA 297
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 91 ProGlnProGlnProLysProGlnProGlnProGlnProGlnProGlnProLys 110
    |||||:::|||||:::|||||:::|||||:::|||||
QY 298 CCGGAAACCGGAAAGTACC 315
    |||||:::|||||:::|||||:::|||||:::|||||
Db 111 ProGlnProGlnCysThr 116
    |||||:::|||||:::|||||:::|||||:::|||||

RESULT 4
TSHUP4
Thrombospondin 4 precursor - human
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 11-Aug-1995 #text_change 13-Aug-1999
C:Accession: A55710; S36069
R:Lawler, J.; McHenry, K.; Duquette, M.; Derick, L.
J. Biol. Chem. 270, 2809-2814, 1995
A:Title: Characterization of human thrombospondin-4.
A:Reference number: A55710; MUID:95155352; PMID:7852353
A:Accession: A55710
A:Molecule type: mRNA
A:Residues: 1-961 <LAN>
A:Cross-references: EMBL:Z19585; NID:9311625; PIDN:CAAT9635.1; PID:9311626
A>Note: authors translated the codon GTG for residue 616 as Ser
C:Genetics:
A:Gene: GDB:THB94
A:Cross-references: GDB:463011; OMIM:600715
A:Map position: 1q21-1q23
A:Complex: homotrimer, disulfide linked
C:Function:
A:Description: participates in cell migration and adhesion, and in platelet aggregation
C:Superfamily: thrombospondin 3; EGF homology
C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; cell adhe
F:12-21/Domain: signal sequence #status predicted <SIG>
F:22-961/Product: thrombospondin 4 #status predicted <MAT>
F:330-362/Domain: EGF homology <EGF1>
F:330-362/Domain: EGF homology <EGF>
F:303/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
F:343/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F:612,941/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 4 04e-05 Length: 961
Score: 145.50 Matches: 30
Percent Similarity: 57.14% Conservative: 10
Best Local Similarity: 42.86% Mismatches: 25

```







A26036  
procyelic acidic repetitive protein A-beta - Trypanosoma brucei  
N/Alternate names: Insect-stage-specific protein  
C/Species: Trypanosoma brucei  
C/Date: 05-Oct-1988 #sequence\_revision 05-Oct-1988 #text\_change 13-Aug-1999  
A/Accession: A26036; S21540  
R/Roditi, I.; Carrington, M.; Turner, M.  
Nature 325, 272-274, 1987  
A>Title: Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*  
A/Reference number: A26036; MUID:87115776; PMID:3808022  
A/Accession: A26036  
A/Molecule type: mRNA  
A/Residues: 1-129 <ROD>  
A/Cross-references: GB:X04814; NID:g10510; PIDN:CAA28503.1; PID:g10511  
R/Vijayarathay, S.; Ernest, I.; Itzhaki, J.; Sherman, D.; Mowat, M.R.; Michels, P.A.M. submitted to the EMBL Data Library, April 1990  
A/Description: The genes encoding fructose biphosphate aldolase in trypanosoma brucei  
A/Reference number: S21538  
A/Accession: S21540  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-129 <VID>  
A/Cross-references: EMBL:X52584; NID:g10480; PIDN:CAA36815.1; PID:g10482  
C/Superfamily: procyelic acidic repetitive protein  
P:63-106/Region: 2-residue repeats (E-F)  
  
Alignment Scores:  
Pred. No.:           0.0112           Length:           129  
Score:               119.50           Matches:          25  
Percent Similarity:   58.82%          Conservative:     15  
Best Local Similarity: 36.76%          Mismatches:      16  
Query Match:           17.78%          Indels:           12  
DB:                    2               Gaps:             3  
  
US-09-696-872-24 (1-372) x A26036 (1-129)  
  
QY           108 GGGCCCGAGATGCTTCGGGAACGCAGAAACAACCAGCGCTGCAGCATCGCGGA 167  
             ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db           30 GlyProGluAsp-----LysGlyLeuThrLySGlyGlyLys----- 43  
  
QY           168 CTGCGTGGGGGAGCAAGTCAGGAGATCATCGTCTGTAAAAAACGCGTAGTGAGTGA 227  
             44 -----GlyGlyLysGlyThrLySValSerAspAPABPTThrAnglyThr As 59  
  
QY           228 CGCGTGGCGGCGCGAGCCGAGCCGAAACCGAGCCGCGAGCGCGAGCCGAGCAAC 287  
             59 pPro---AspProGluInProGluInProGluInProGluInProGluInProGluInPr 78  
  
QY           288 GCAGCCGAAACCGGAACCGGAA 309  
             :::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||:::|||  
Db           78 oGluProGluProGluProGlu 85  
  
RESULT 12  
T24470  
hypothetical protein T04F8.8 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
A/Accession: T24470  
R/Lennard, N.  
submitted to the EMBL Data Library, November 1995  
A/Reference number: Z19895  
A/Accession: T24470  
A/Status: preliminary; translated from GB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-165 <MIT>  
A/Cross-references: EMBL:Z66565; PIDN:CAA91483.1; GSPDB:GN00028; CESP:T04F8.8  
A/Experimental source: clone T04F8  
C/Genetics:  
A/Gene: CESP:T04F8.8  
A/Map position: X  
A/introns: 18/2, 63/1  
  
Alignment Scores:

[illegible]



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 26, 2003, 16:07:14 ; Search time 17.5 Seconds  
(without alignments)  
1999.307 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 672  
Sequence: 1 aaagctaccatcggaagta.....aattccatcatcgatcgag 372

Scoring table:

	BLOSUM62
Xgapop 10.0 , Xgapext 0.5	
Xgapop 10.0 , Xgapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO\_epool/US09696872/runat\_26082003.150542.8719/app\_query.fasta.1.519  
-DB=swissprot 41 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09696872 @CGN 1.1.30 @runat 26082003.150542.8719 -NCP=6 -ICPU=3  
-NO\_MMAP -LARGESOURCE -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=110 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262.5	39.1	757	1 COMP_HUMAN	P49747 homo sapien
2	234	34.8	755	1 COMP_RAT	P35444 rattus norv
3	146.5	21.8	955	1 TSP4_XENLA	Q05441 xenopus lae
4	145.5	21.7	961	1 TSP4_HUMAN	P35443 homo sapien
5	144.5	21.5	980	1 TSP4_RAT	P49744 rattus norv
6	133	19.8	339	1 CSP_PLAAB	P06591 plasmodium
7	133	19.8	347	1 CSP_PLAAB	P23093 plasmodium
8	132	19.6	3164	1 TEGU_HSV1	P10220 herpes simp
9	124	18.5	220	1 NOJ3_MOUSE	Q96100 mus musculi
10	122	18.2	283	1 SFR4_MOUSE	Q06701 mus musculi
11	119.5	17.8	129	1 PARB_TRYBB	P05791 trypanosoma
12	117	17.4	285	1 TONB_HELPY	O25899 helicobacte
13	116.5	17.3	115	1 PARX_TRYBB	P14043 trypanosoma
14	116.5	17.3	143	1 PAR1_TRYBB	P08469 trypanosoma
15	116.5	17.3	143	1 PARC_TRYBB	Q05084 trypanosoma
16	116	16.5	194	1 KRCB_HUMAN	Q75690 homo sapien
17	115	17.1	280	1 TONB_HELPY	Q92jP4 helicobacte
18	113.5	16.9	221	1 NOJ3_RAT	Q62881 rattus norv

19	112.5	16.7	839	1 TRA9_MAIZE	P03010 zea mays (m
20	112.5	16.7	1569	1 YPUA_ECOLI	P52143 escherichia
21	111	16.5	806	1 TRAI_MAIZE	P08770 zea mays (m
22	109.5	16.3	221	1 SFR2_CHICK	P30352 gallus gall
23	108	16.1	722	1 Z219_HUMAN	Q9P2Y4 homo sapien
24	107	15.9	503	1 VE2_HPV21	P50767 human papil
25	107	15.9	613	1 PKPA_PHYBL	Q01577 phycomyces
26	106	15.8	955	1 T150_HUMAN	Q9Y2W1 homo sapien
27	105.5	15.0	131	1 CHHB_BOMMO	P05688 bombyx mori
28	105	15.6	238	1 SFR7_HUMAN	Q16629 homo sapien
29	105	15.6	243	1 TONB_KLEBN	P45610 klebsiella
30	104.5	15.6	221	1 SFR2_HUMAN	Q01130 homo sapien
31	104.5	15.6	221	1 UZR2_MOUSE	O62093 mus musculi
32	104.5	15.6	482	1 UZR2_HUMAN	Q15696 homo sapien
33	104.5	15.6	1205	1 S122_MOUSE	P55012 mus musculi
34	103.5	15.4	375	1 SRS5_DROME	P26686 drosophila
35	103.5	15.4	483	1 VE2_HPV14	P36783 human papil
36	103.5	15.4	493	1 VE2_HPV19	P36786 human papil
37	103.5	15.4	1493	1 M3K1_RAT	Q62925 rattus norv
38	103	15.3	269	1 SFR5_RAT	Q09167 rattus norv
39	102.5	15.3	494	1 SFR4_HUMAN	Q08170 homo sapien
40	102.5	15.3	497	1 VE2_HPV20	P50766 human papil
41	102	15.2	449	1 APG_BRANA	P40603 brassica na
42	101	15.0	270	1 SFR5_MOUSE	O51368 pseudomnas
43	101	15.0	342	1 TONB_PSEAE	P09516 barley yell
44	101	15.0	450	1 V50K_BYDVP	
45	101	15.0	565	1 MOT8_MOUSE	O70324 mus musculi

#### ALIGNMENTS

##### RESULT 1

ID	COMP_HUMAN	STANDARD	PRT	757 AA.
AC	P49747; Q16388; Q16389;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Cartilage oligomeric matrix protein precursor (COMP).			
GN	COMP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]	SEQUENCE FROM N.A.		
RP	TISSUE=Cartilage;			
RC	MEDLINE=95229140; PubMed=7713493;			
RA	Newton G., Meremowicz S., Morton C.C., Copeland N.G.,			
RA	Gilbert D.J., Jenkins N.A., Lawler J.,			
RT	"Characterization of human and mouse cartilage oligomeric matrix			
RT	protein.";			
RL	Genomics 24:435-439(1994).			
[2]				
RP	VARIANTS PSACH SER-459 DEL; TYR-468 AND TYR-472.			
RX	MEDLINE=95400301; PubMed=7670471.			
RA	Hecht J.T., Nelson L.D., Crowder E., Wang Y., Elder F.F.B.,			
RA	Harrison W.R., Francomano C.A., Prange C.K., Lennon G.G., Deere M.,			
RA	Lawler J.;			
RT	"Mutations in exon 17B of cartilage oligomeric matrix protein (COMP)			
RT	cause pseudochondroplasia.";			
RL	Nat. Genet. 10:325-329(1995).			
[3]				
RP	VARIANTS MED TYR-342, AND VARIANT PSACH ARG-328.			
RX	MEDLINE=95400302; PubMed=7670472;			
RA	Briggs M.D., Hoffman S.M.G., King L.M., Olsen A.S., Mohrenweiser H.,			
RA	Leroy J.G., Mortier G.R., Rimojn D.L., Lachman R.S., Gaines E.S.,			
RA	Cekleniak J.A., Knowlton R.G., Cohn D.H.;			
RT	"Pseudochondroplasia and multiple epiphyseal dysplasia due to			
RT	mutations in the cartilage oligomeric matrix protein gene.";			
RL	Nat. Genet. 10:330-336(1995).			
[4]				
RP	VARIANT MED LYS-523.			

RX MEDLINE=97173141; PubMed=9021009;  
 RA Balle R., Briggs M.D., Cohn D.H., Knowlton R.G., Beighton P.H.,  
 RA Ramesar R.S.;  
 RT "Multiple epiphyseal dysplasia, ribbing type: a novel point mutation  
 RT in the COMP gene in a South African family";  
 RL Am. J. Med. Genet. 68:396-400(1997).  
 RN (5)  
 RP VARIANT MED SER-371, AND VARIANT PSACH 513-VAL-LYS-516 DEL.  
 RX MEDLINE=97327574; PubMed=9184241;  
 RA Susic S., McGrovy J., Ahler J., Cole W.G.;  
 RT "Multiple epiphyseal dysplasia and pseudoachondroplasia due to novel  
 RT mutations in the calmodulin-like repeats of cartilage oligomeric  
 RT matrix protein";  
 RL Clin. Genet. 51:219-224(1997).  
 RN (6)  
 RP VARIANTS PSACH AND MED.  
 RX MEDLINE=98130533; PubMed=9463320;  
 RA Briggs M.D., Mortier G.R., Cole W.G., King L.M., Golik S.S.,  
 RA Bonneventure J., Nuytink L., de Paepe A., Leroy J.G., Bleecker L.,  
 RA Lipson M., Wilcox W.R., Lachman R.S., Rimojn D.L., Knowlton R.G.,  
 RA Cohn D.H.;  
 RT "Diverse mutations in the gene for cartilage oligomeric matrix protein  
 RT in the pseudoachondroplasia-multiple epiphyseal dysplasia disease  
 RT spectrum";  
 RL Am. J. Hum. Genet. 62:311-319(1998).  
 RN (7)  
 RP VARIANTS PSACH AND MED.  
 RX MEDLINE=99118868; PubMed=9921895;  
 RA Ikegawa S., Ohashi H., Nishimura G., Kim K.C., Sannohe A.,  
 RA Kimizuka M., Fukushima Y., Nagai T., Nakamura Y.,  
 RT "Novel and recurrent COMP (cartilage oligomeric matrix protein)  
 RT mutations in pseudoachondroplasia and multiple epiphyseal dysplasia";  
 RL Hum. Genet. 103:633-638(1998).  
 RN (8)  
 RP VARIANTS PSACH AND MED.  
 RX MEDLINE=98112405; PubMed=9452026;  
 RA Loughlin J., Irven C., Mustafa Z., Briggs M.D., Carr A., Lynch S.A.,  
 RA Knowlton R.G., Cohn D.H., Sykes B.;  
 RT "Identification of five novel mutations in cartilage oligomeric  
 RT matrix protein gene in pseudoachondroplasia and multiple epiphyseal  
 RT dysplasia";  
 RL Hum. Mutat. Suppl. 1:S10-S17(1998).  
 RN (9)  
 RP VARIANT PSACH GLY-482.  
 RX MEDLINE=98112442; PubMed=9452063;  
 RA Susic S., Ahler J., Cole W.G.;  
 RT "Pseudoachondroplasia due to the substitution of the highly conserved  
 RT Asp82 by Gly in the seventh calmodulin-like repeat of cartilage  
 RT oligomeric matrix protein";  
 RL Hum. Mutat. Suppl. 1:S125-S127(1998).  
 CC -1- SUBUNIT: PENTAMER; DISULFIDE-LINKED.  
 CC -1- DISEASE: DEFECTS IN COMP ARE THE CAUSE OF PSEUDOACHONDROPLASIA  
 CC (PEACH) AND MULTIPLE EPIPHYSEAL DYSPLASIA (MED OR EDM1), WHICH ARE  
 CC DOMINANTLY INHERITED CHONDRODYSPLASIAS CHARACTERIZED BY SHORT  
 CC STATURE AND EARLY-ONSET OSTEOARTHRITIS. MED IS BROADLY CATEGORIZED  
 CC INTO THE MORE SEVERE FAIRBANK AND THE Milder RIBBING TYPES. PSACH  
 CC IS MORE SEVERE AND IS RECOGNIZED IN EARLY CHILDHOOD.  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1- SIMILARITY: Contains 4 EGF-like domains.  
 CC -1- SIMILARITY: Contains 7 TSP type-3 domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sb.ch/announce/>  
 CC or send an email to [license@isb-sb.ch](mailto:license@isb-sb.ch)).  
 CC -----  
 DR EMBL: L32137; AAA57253.1; -  
 DR EMBL: S79499; AAB35289.1; -  
 DR EMBL: S79500; AAB35270.1; -  
 DR HGSP: P35444; 1VDF.

DR	Genew; HGNC:2227; COMP.	
DR	MIM; 600310; -.	
DR	MIM; 132400; -.	
DR	MIM; 177170; -.	
DR	GO; GO:0005578; C:extracellular matrix; TAS.	
DR	GO; GO:0005509; P:calcium ion binding activity; TAS.	
DR	GO; GO:0005501; P:extracellular matrix structural constituent; TAS.	
DR	GO; GO:0007397; P:histogenesis and organogenesis; TAS.	
DR	GO; GO:001501; P:skeletal development; TAS.	
DR	InterPro; IPRO01881; EGF_Ca.	
DR	InterPro; IPRO06209; EGF_like.	
DR	InterPro; IPRO03367; tsp_3.	
DR	Pfam; PF00008; EGF; 2.	
DR	Pfam; PF02412; tsp_3; 11.	
DR	SMART; SM00179; EGF_Ca; 2.	
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.	
DR	PROSITE; PS01186; EGF_2; 1.	
DR	PROSITE; PS01187; EGF_Ca; 2.	
KW	Signal; Cell adhesion; Calcium-binding; Repeat; EGF-like domain.	
KW	Signal; Disease mutation.	
FT	SIGNAL	1 20
FT	CHAIN	21 757
FT	DOMAIN	22 86
FT	DOMAIN	87 126
FT	DOMAIN	127 179
FT	DOMAIN	180 222
FT	DOMAIN	225 267
FT	DOMAIN	297 332
FT	DOMAIN	333 355
FT	DOMAIN	356 391
FT	DOMAIN	392 414
FT	DOMAIN	415 452
FT	DOMAIN	453 488
FT	DOMAIN	489 524
FT	DOMAIN	525 757
FT	DISUPEID	69 69
FT	DISUPEID	72 72
FT	DISUPEID	91 102
FT	DISUPEID	96 111
FT	DISUPEID	114 125
FT	DISUPEID	131 142
FT	DISUPEID	136 151
FT	DISUPEID	154 178
FT	DISUPEID	184 197
FT	DISUPEID	191 206
FT	DISUPEID	209 221
FT	DISUPEID	229 243
FT	DISUPEID	237 253
FT	DISUPEID	255 266
FT	DISUPEID	282 317
FT	DISUPEID	292 382
FT	DISUPEID	328 348
FT	DISUPEID	351 371
FT	DISUPEID	387 407
FT	DISUPEID	410 430
FT	DISUPEID	448 468
FT	DISUPEID	484 504
FT	DISUPEID	520 741
FT	CARBOHYD	121 121
FT	CARBOHYD	742 742
FT	CARBOHYD	367 369
FT	SITE	290 290
FT	VARIANT	299 299
FT	VARIANT	328 328
FT	VARIANT	342 342
FT	VARIANT	349 349
FT	VARIANT	361 361
FT	VARIANT	361 361

FT VARIANT 361 361 D -> Y (IN MED).  
FT /FTID=VAR\_007620.  
FT VARIANT 367 368 MISSING (IN MED).  
FT /FTID=VAR\_007621.  
FT VARIANT 371 371 C -> S (IN MED) FAIRBANK TYPE).  
FT /FTID=VAR\_007622.  
FT VARIANT 372 372 MISSING (IN PSACH).  
FT /FTID=VAR\_007623.  
FT VARIANT 374 374 MISSING (IN PSACH; MILD FORM).  
FT /FTID=VAR\_007624.  
FT VARIANT 387 387 C -> G (IN PSACH; MILD FORM).  
FT /FTID=VAR\_007625.  
FT VARIANT 391 394 PMSD -> V (IN PSACH).  
FT /FTID=VAR\_007626.

Alignment Scores:  
Pred. No.: 2.1e-14 Length: 757  
Score: 262.50 Matches: 57  
Percent Similarity: 75.61% Conservative: 5  
Best Local Similarity: 69.51% Mismatches: 9  
Query Match: 39.06% Indels: 11  
DB: 1 Gaps: 2

US-09-696-872-24 (1-372) x COMP\_HUMAN (1-757)

OY 34 TTGCTCGCCCTTGGCGAGCTGTGACGCGTGCAGCAAAAAGATCCAGCGTGGTGAGAC 93  
Db 10 LeuLeuThrLeuAlaAlaLeuGlyAlaSerGlyGlnGlnSerProLeuGly----- 27  
OY 94 TGTGTTCAGACCTGTGGGCGCGAGATGCTTGGGAACTGCAGAAACACGCGCGCTG 153  
Db 28 -----SerAspLeuGlyProGlnMetLeuArgGlnLeuGlnGlnThrAsnAlaLeu 45  
OY 154 CAGGACGTCGGGAGCTGCTGTGGGACGAGTGCAGGAGGATGATCTTCTGAAAAACG 213  
Db 46 GlnAspValaGAspTrpLeuArgGlnGlnValaGlnGlnIleThrPheLeuLysAsnThr 65  
OY 214 GTGATGAGTGTGACGCGCGCGCGCGAG-----CCGACGCG 252  
Db 66 ValMetGluCysAspAlaCysGlyMetGlnGlnSerValaArgThrGlyLeuProSerVal 85  
OY 253 AAACCG 258  
Db 86 ArgPro 87  
RESULT 2  
COMP\_RAT STANDARD; PRT; 755 AA.  
AC P35444;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Cartilage oligomeric matrix protein precursor (COMP).  
GN COMP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN (1)  
RN SEQUENCE FROM N.A.  
RC TISSUE=Cartilage;  
RX MEDLINE=93054522; PubMed=1429587;  
RX Oldberg A., Antonsson P., Lindblom K., Heinegaard D.;  
RT "COMP (cartilage oligomeric matrix protein) is structurally related  
RL to the thrombospondins.";  
RL J. Biol. Chem. 267:22346-22350(1992).  
RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 27-72.  
RX MEDLINE=97020114; PubMed=8864111;  
RA Malaebhevi V.N., Kammerer R.A., Efimov V.P., Schultze T.,  
RA Engel J.;  
RT "The crystal structure of a five-stranded coiled coil in COMP: a  
RT prototype ion channel?";

RL Science 274:761-765(1996).  
CC -1- SUBUNIT: PENTAMER; DISULFIDE-LINKED.  
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
CC -1- SIMILARITY: Contains 4 EGF-like domains.  
CC -1- SIMILARITY: Contains 7 TSP type-3 domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by, and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X72914; CA51419.1; -.  
DR PIR: A44315; A44315.  
DR PDB: 1VDF; 08-OCT-97.  
DR PDB: 1FBM; 09-AUG-00.  
DR InterPro: IPR001881; EGF\_Ca.  
DR InterPro: IPR006209; EGF\_Like.  
DR InterPro: IPR003367; tsp\_3.  
DR Pfam: PF00008; EGF; 1.  
DR Pfam: PF02412; tsp\_3; 9.  
DR SMART: SM00179; EGF\_CA; 2.  
DR PROSITE: PS00022; EGF\_1; FALSE\_NEG.  
DR PROSITE: PS01186; EGF\_2; 1.  
DR PROSITE: PS01187; EGF\_CA; 2.  
KM Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;  
KW Signal; 3D-structure.  
FT SIGNAL 1 19  
FT CHAIN 20 755  
FT DOMAIN 21 84  
FT DOMAIN 85 124  
FT DOMAIN 125 177  
FT DOMAIN 178 220  
FT DOMAIN 223 265  
FT DOMAIN 295 330  
FT DOMAIN 331 353  
FT DOMAIN 354 389  
FT DOMAIN 390 412  
FT DOMAIN 413 450  
FT DOMAIN 451 486  
FT DOMAIN 487 522  
FT DOMAIN 523 755  
FT DISULFID 68 68  
FT DISULFID 71 71  
FT DISULFID 89 100  
FT DISULFID 94 109  
FT DISULFID 112 123  
FT DISULFID 129 140  
FT DISULFID 134 149  
FT DISULFID 152 176  
FT DISULFID 182 195  
FT DISULFID 189 204  
FT DISULFID 207 219  
FT DISULFID 227 241  
FT DISULFID 235 251  
FT DISULFID 253 264  
FT DISULFID 280 285  
FT DISULFID 290 310  
FT DISULFID 326 346  
FT DISULFID 349 369  
FT DISULFID 385 405  
FT DISULFID 408 428  
FT DISULFID 446 466  
FT DISULFID 482 502  
FT DISULFID 518 739  
FT CARBOHYD 119 119  
FT CARBOHYD 740 740  
FT CARBOHYD 740 740  
FT HELIX 30 66  
FT TURN 67 67  
FT TURN 69 71  
FT HELIX 71 71  
SQ SEQUENCE 755 AA; 82663 MW; AB48888FE093C598 CRC64;

```

Alignment Scores:
Pred. No.: 4.76e-12 Length: 755
Score: 234.00 Matches: 51
Percent Similarity: 72.84% Conservative: 8
Best Local Similarity: 62.96% Mismatch: 18
Query Match: 34.82% Indel: 4
DB: 1 Gaps: 1

US-09-696-872-24 (1-372) x COMP_RAT (1-755)

QY 34 TTGCTCGCCCTTGGCGGAGTGTGCAGCGCGTCCAAAAAGATTCAGCGCTGGTGAGAC 93
   :::::::::::::::::::::
Db 9 ValLeuAlaLeuAlaLeuAlaArgAlaThrGlyGlnGlyGlnIleProLeuGlyGly--- 27
   ValLeuAlaLeuAlaLeuAlaArgAlaThrGlyGlnGlyGlnIleProLeuGlyGly--- 27

QY 94 TGTGTTGACGACCTGGGGCCCGACAGTGTGGGAACTGGACGAAACCAACGCGCGCTG 153
   -----AspLeuAlaProGlnMetLeuArgGlnLeuGlnGlnIleThrAsnAlaLeu 44
   -----AspLeuAlaProGlnMetLeuArgGlnLeuGlnGlnIleThrAsnAlaLeu 44

QY 154 CAGGACGTGCGGAGCTGGCTGCGGACGAGGTACAGAGATCACGTTCTGAAAAACAG 213
   |||||:::
Db 45 GlnsppValaArgGlnLeuLeuAlaArgAlaThrGlyGlnGlyGlnIleThrPheLeuLysAsnThr 64
   |||||:::

QY 214 GTGATGAGTGTGACGCGCTGCGGCGCGGACCGGACCGGAAACCGACCGCAGCCGAG 273
   |||||
Db 65 ValMetGlnGlyAspAlaGlyGlyMetGlnProAlaArgThrProGlyLeuSerValaArg 84
   |||||

QY 274 CCG 276
   |||
Db 85 Pro 85

RESULT 3
TSP4_XENULA STANDARD; PRT; 955 AA.
ID TSP4_XENULA STANDARD; PRT; 955 AA.
AC Q06441;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thrombospondin 4 precursor.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93163109; PubMed=8432726;
RA Lawler J., Duquette M., Whitaker C.A., Adams J.C., McHenry K.,
RA Desjardine D.W.;
RT "Identification and characterization of thrombospondin-4, a new
RT member of the thrombospondin gene family."
RL J. Cell Biol. 120:1059-1067(1993).
CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC LAMININ AND TYPE V COLLAGEN. MAY PARTICIPATE IN THE GENESIS AND
CC FUNCTION OF CARDIAC AND SKELETAL MUSCLE.
CC -1- SUBUNIT: Homotrimer; disulfide-linked.
CC -1- DEVELOPMENTAL STAGE: INITIAL EXPRESSION DURING NEURULATION.
CC INCREASE DURING TAILBUD STAGES BUT DECREASE BY THE FEEDING TADPOLE
CC STAGE.
CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC -1- SIMILARITY: Contains 4 EGF-like domains.
CC -1- SIMILARITY: Contains 7 TSP type-3 domains.
CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC

```

DR	EMBL; Z19091; CAA79518.1; -.		
DR	PIR; A45441; A45441.		
DR	HSP: P35444; 1VDF.		
DR	InterPro; IPR001881; EGF_Ca.		
DR	InterPro; IPR006209; EGF_like.		
DR	InterPro; IPR003367; tsp_3.		
DR	InterPro; IPR003129; TSPN.		
DR	Pfam; PF00008; EGF; 2.		
DR	Pfam; PF02412; tsp_3; 9.		
DR	Pfam; PF02210; TSPN; 1.		
DR	SMART; SM00179; EGF_Ca; 2.		
DR	SMART; SM00210; TSPN; 1.		
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.		
DR	PROSITE; PS01186; EGF_2; 1.		
DR	PROSITE; PS01187; EGF_Ca; 2.		
KW	Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain; Signal.		
FT	Signal.	24	POTENTIAL.
FT	CHAIN	1	THROMBOSPONDIN 4.
FT	DOMAIN	25	N-TERMINAL.
FT	DOMAIN	25	EGF-LIKE 1.
FT	DOMAIN	281	EGF-LIKE 2.
FT	DOMAIN	321	EGF-LIKE 3.
FT	DOMAIN	374	EGF-LIKE 4.
FT	DOMAIN	418	EGF-LIKE 5.
FT	DOMAIN	489	EGF-LIKE 6.
FT	DOMAIN	525	EGF-LIKE 7.
FT	DOMAIN	548	EGF-LIKE 8.
FT	DOMAIN	584	EGF-LIKE 9.
FT	DOMAIN	607	EGF-LIKE 10.
FT	DOMAIN	645	EGF-LIKE 11.
FT	DOMAIN	685	EGF-LIKE 12.
FT	DOMAIN	714	EGF-LIKE 13.
FT	DOMAIN	756	EGF-LIKE 14.
FT	DOMAIN	785	EGF-LIKE 15.
FT	DOMAIN	825	EGF-LIKE 16.
FT	DOMAIN	865	EGF-LIKE 17.
FT	DOMAIN	905	EGF-LIKE 18.
FT	DOMAIN	945	EGF-LIKE 19.
FT	DOMAIN	985	EGF-LIKE 20.
FT	DOMAIN	1025	EGF-LIKE 21.
FT	DOMAIN	1065	EGF-LIKE 22.
FT	DOMAIN	1105	EGF-LIKE 23.
FT	DOMAIN	1145	EGF-LIKE 24.
FT	DOMAIN	1185	EGF-LIKE 25.
FT	DOMAIN	1225	EGF-LIKE 26.
FT	DOMAIN	1265	EGF-LIKE 27.
FT	DOMAIN	1305	EGF-LIKE 28.
FT	DOMAIN	1345	EGF-LIKE 29.
FT	DOMAIN	1385	EGF-LIKE 30.
FT	DOMAIN	1425	EGF-LIKE 31.
FT	DOMAIN	1465	EGF-LIKE 32.
FT	DOMAIN	1505	EGF-LIKE 33.
FT	DOMAIN	1545	EGF-LIKE 34.
FT	DOMAIN	1585	EGF-LIKE 35.
FT	DOMAIN	1625	EGF-LIKE 36.
FT	DOMAIN	1665	EGF-LIKE 37.
FT	DOMAIN	1705	EGF-LIKE 38.
FT	DOMAIN	1745	EGF-LIKE 39.
FT	DOMAIN	1785	EGF-LIKE 40.
FT	DOMAIN	1825	EGF-LIKE 41.
FT	DOMAIN	1865	EGF-LIKE 42.
FT	DOMAIN	1905	EGF-LIKE 43.
FT	DOMAIN	1945	EGF-LIKE 44.
FT	DOMAIN	1985	EGF-LIKE 45.
FT	DOMAIN	2025	EGF-LIKE 46.
FT	DOMAIN	2065	EGF-LIKE 47.
FT	DOMAIN	2105	EGF-LIKE 48.
FT	DOMAIN	2145	EGF-LIKE 49.
FT	DOMAIN	2185	EGF-LIKE 50.
FT	DOMAIN	2225	EGF-LIKE 51.
FT	DOMAIN	2265	EGF-LIKE 52.
FT	DOMAIN	2305	EGF-LIKE 53.
FT	DOMAIN	2345	EGF-LIKE 54.
FT	DOMAIN	2385	EGF-LIKE 55.
FT	DOMAIN	2425	EGF-LIKE 56.
FT	DOMAIN	2465	EGF-LIKE 57.
FT	DOMAIN	2505	EGF-LIKE 58.
FT	DOMAIN	2545	EGF-LIKE 59.
FT	DOMAIN	2585	EGF-LIKE 60.
FT	DOMAIN	2625	EGF-LIKE 61.
FT	DOMAIN	2665	EGF-LIKE 62.
FT	DOMAIN	2705	EGF-LIKE 63.
FT	DOMAIN	2745	EGF-LIKE 64.
FT	DOMAIN	2785	EGF-LIKE 65.
FT	DOMAIN	2825	EGF-LIKE 66.
FT	DOMAIN	2865	EGF-LIKE 67.
FT	DOMAIN	2905	EGF-LIKE 68.
FT	DOMAIN	2945	EGF-LIKE 69.
FT	DOMAIN	2985	EGF-LIKE 70.
FT	DOMAIN	3025	EGF-LIKE 71.
FT	DOMAIN	3065	EGF-LIKE 72.
FT	DOMAIN	3105	EGF-LIKE 73.
FT	DOMAIN	3145	EGF-LIKE 74.
FT	DOMAIN	3185	EGF-LIKE 75.
FT	DOMAIN	3225	EGF-LIKE 76.
FT	DOMAIN	3265	EGF-LIKE 77.
FT	DOMAIN	3305	EGF-LIKE 78.
FT	DOMAIN	3345	EGF-LIKE 79.
FT	DOMAIN	3385	EGF-LIKE 80.
FT	DOMAIN	3425	EGF-LIKE 81.
FT	DOMAIN	3465	EGF-LIKE 82.
FT	DOMAIN	3505	EGF-LIKE 83.
FT	DOMAIN	3545	EGF-LIKE 84.
FT	DOMAIN	3585	EGF-LIKE 85.
FT	DOMAIN	3625	EGF-LIKE 86.
FT	DOMAIN	3665	EGF-LIKE 87.
FT	DOMAIN	3705	EGF-LIKE 88.
FT	DOMAIN	3745	EGF-LIKE 89.
FT	DOMAIN	3785	EGF-LIKE 90.
FT	DOMAIN	3825	EGF-LIKE 91.
FT	DOMAIN	3865	EGF-LIKE 92.

QY	108	GGCGCCGAGATGCTGGGAACTGCAGAGAAACCAACCGCGGCGTGCAGGACGTGCGGGA	167
Db	217	ISerArgGlnIleuIleGlyGlnIleThrGlnMetCasnIleMetLeuGlyGlnLeuArgAs	237
QY	168	CTGCCTGCGCAGCAGGTCTCAGGAGATCACGTTCCTGAAAAACAGGTGATGAGTGTGA	227
		:::	
Db	237	pValMetArgGlnGlnValValysGlnThrMetPheLeuArgGlnThrIleAlaGlnCysGln	257
QY	228	CGCGTGCGGGCGCAGCCGCAG--CCGAAACCGCAGCCGCGCAGC-----	273
Db	257	nAlaCysGlnLeuGlyProAlaPhePheProLeuProThrIysValProGlnArgLeuAlaTh	277
QY	274	-----CCGAGCGCGCAACCG-----	288
Db	277	rThrThrProProIysProIleArgCysAspAlaThrSerCysPheArgIleValArgCysIle	297
QY	289	-----CAGCCGAAACCGGAAACCGGAGGT---ACCGATCA-----	321
Db	297	eAspThrGlnIleGlyGlnPheGlnCysGlyProCysPheProGlnIleGlyrThrGlyAsnGlyVa	317
QY	322	-----TCGAAAAAAGATGAGTGTGTAG	343
Db	317	IleCysThrAspValaIleGlu-CysArg	326
RESULT 4			
TSp4 HUMAN			
ID	TSp4 HUMAN	STANDARD;	PRT; 961 AA.
AC	P35443;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Thrombospondin 4 precursor.		
GN	THBS4 OR TSP4.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Heart;		
RX	MEDLINE=93353522; PubMed=8350346;		
RA	Lawler J., Duquette M., Urry L., McHenry K., Smith T.F.;		
RT	"The evolution of the thrombospondin gene family.";		
RL	J. Mol. Evol. 36:509-516(1993).		
CC	-1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND		
CC	CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBONECTIN,		
CC	LAMININ AND TYPE V COLLAGEN.		
CC	-1- SUBUNIT: Homotrimer; disulfide-linked.		
CC	-1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.		
CC	-1- SIMILARITY: Contains 4 EGF-like domains.		
CC	-1- SIMILARITY: Contains 7 TSP type-3 domains.		
CC	-1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation on		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		
DR	EMBL; Z19585; CAA79635.1; -		
DR	PIR; A55710; TSHU4.		
DR	HSSP; P35444; 1YDF.		
DR	Genew; HGNC:11788; THBS4.		
DR	MIM; 600715; -		
DR	GO; GO:0005578; C:extracellular matrix; TAS.		
DR	GO; GO:0005509; F:calcium ion binding activity; TAS.		
DR	GO; GO:0008201; F:heparin binding activity; TAS.		
DR	GO; GO:0006930; P:substrate-bound cell migration, cell extension; TAS.		
DR	InterPro; IPR001881; EGF_Ca		
DR	InterPro; IPR006209; EGF-like		
DR	InterPro; IPR003367; tsp_3.		

DR	InterPro:IPR003129; TSPN.
DR	Pfam; PF00008; EGF_2.
DR	Pfam; PF02412; TSP_3; 9.
DR	Pfam; PF02210; TSPN; 1.
DR	SMART; SMO0179; EGF_CA; 2.
DR	SMART; SMO0210; TSPN; 1.
DR	PROSITE; PS00022; EGF_1; FALSE_NEG.
DR	PROSITE; PS01186; EGF_2; 1.
DR	PROSITE; PS01187; EGF_CA; 2.
KW	Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain; Signal.
FT	SIGNAL 1 21 POTENTIAL.
FT	CHAIN 22 961 THROMBOSPONDIN 4.
FT	DOMAIN 22 285 N-TERMINAL.
FT	DOMAIN 286 325 EGF-LIKE 1.
FT	DOMAIN 326 378 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 379 419 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 420 462 EGF-LIKE 4.
FT	DOMAIN 492 527 TSP TYPE-3 1.
FT	DOMAIN 528 550 TSP TYPE-3 2.
FT	DOMAIN 551 586 TSP TYPE-3 3.
FT	DOMAIN 587 609 TSP TYPE-3 4.
FT	DOMAIN 610 647 TSP TYPE-3 5.
FT	DOMAIN 648 687 TSP TYPE-3 6.
FT	DOMAIN 688 723 TSP TYPE-3 7.
FT	DOMAIN 724 961 C-TERMINAL.
FT	SITE 562 564 CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID 258 258 INTERCHAIN (PROBABLE).
FT	DISULFID 261 261 INTERCHAIN (PROBABLE).
FT	DISULFID 290 301 BY SIMILARITY.
FT	DISULFID 295 310 BY SIMILARITY.
FT	DISULFID 313 324 BY SIMILARITY.
FT	DISULFID 330 341 BY SIMILARITY.
FT	DISULFID 335 350 BY SIMILARITY.
FT	DISULFID 353 377 BY SIMILARITY.
FT	DISULFID 383 394 BY SIMILARITY.
FT	DISULFID 388 403 BY SIMILARITY.
FT	DISULFID 406 418 BY SIMILARITY.
FT	DISULFID 424 438 BY SIMILARITY.
FT	DISULFID 432 448 BY SIMILARITY.
FT	DISULFID 450 461 BY SIMILARITY.
FT	DISULFID 477 482 BY SIMILARITY.
FT	DISULFID 487 507 BY SIMILARITY.
FT	DISULFID 523 543 BY SIMILARITY.
FT	DISULFID 546 566 BY SIMILARITY.
FT	DISULFID 582 602 BY SIMILARITY.
FT	DISULFID 605 625 BY SIMILARITY.
FT	DISULFID 643 663 BY SIMILARITY.
FT	DISULFID 683 703 BY SIMILARITY.
FT	DISULFID 719 940 BY SIMILARITY.
FT	CARBOHYD 612 612 N-LINKED (GLCNAC... ) (POTENTIAL).
FT	CARBOHYD 941 941 N-LINKED (GLCNAC... ) (POTENTIAL).
SO	SEQUENCE 961 AA; 105801 MW; 55A4BAF481ABE9DD CRC64;
Alignment Scores:	
Pred. No.:	9, 86e-05 Length: 961
Score:	145.50 Matches: 30
Percent Similarity:	57.14% Conservative: 10
Best Local Similarity:	42.86% Mismatches: 25
Query Match:	21.65% Indels: 5
DB:	Gaps: 2
US-09-696-872-24 (1-372) x TSP4_HUMAN (1-961)	
OY	103 GACCTGAGCCCGCAATGCTTCGGGAAGTGCAGAAACCAACGCCGCGCTGCAAGACGTG 162
Db	218 AspPheAsnArgGlnPheLeucIyImetThrGlnLeuAenGlnLeuLgylVal 237
OY	163 CGGAGTAGCTCCTCGGAGAGCATGCAAGTCAAGTTCCGTAAGAAAACAGTGATGAG 222
Db	238 LysAspLeuLeuArgGlnInValIysGlnThrSerPheLeuArgAenThrIleAlaGlu 257
OY	223 TTGAGACGCTGCGGAGCGGCGG-----CAGCGCAGCGCGAACCAGCGCAGCGCAGCGG 270

Db 258 CysGlnAlaCyGlyProLeuLysPheGlnSerProThrProSerThrValAlaPro 277  
 QY 271 CAGCCGACCCGAAACCGCAGCCGAAACCG 300  
 Db 278 AlaPro---ProAlaProProThrArgPro 286

RESULT 5  
 TSP4\_RAT STANDARD; PRT; 980 AA.  
 AC P49744;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Thrombospondin 4 precursor.  
 GN THB4 OR TSP4 OR TSP-4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 NCBI\_TaxID=10116;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Lewis; TISSUE=Skeletal muscle;  
 RX MEDLINE=96074771; PubMed=7490284;  
 RA Arber S., Caroni P.;  
 RT "Thrombospondin-4, an extracellular matrix protein expressed in the  
 RT developing and adult nervous system promotes neurite outgrowth.";  
 RL J. Cell Biol. 131:1083-1094(1995).  
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
 CC LAMININ AND TYPE V COLLAGEN.  
 CC -1- SUBUNIT: HOMOTRIMER, CROSS-LINKED BY DISULFIDE BONDS (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1- SIMILARITY: Contains 4 EGF-like domains.  
 CC -1- SIMILARITY: Contains 7 TSP type-3 domains.  
 CC -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X89963; CAA62002.1; -.  
 DR HSSP; P35444; 1YDF.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR006209; EGF\_Like.  
 DR InterPro; IPR003367; TSP\_3.  
 DR InterPro; IPR003129; TSPN.  
 DR Pfam; PF000008; EGF; 2.  
 DR Pfam; PF02412; TSP\_3; 9.  
 DR Pfam; PF02210; TSPN; 1.  
 DR SMART; SM00179; EGF\_CA; 2.  
 DR SMART; SM00210; TSPN; 1.  
 DR PROSITE; PS00022; EGF\_1; FALSE\_NEG.  
 DR PROSITE; PS01186; EGF\_2; 1.  
 DR PROSITE; PS01187; EGF\_CA; 2.  
 DR GlycoProtein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain;  
 KM Signal.  
 FT CHAIN 1 39 POTENTIAL.  
 FT CHAIN 40 980 THROMBOSPONDIN 4.  
 FT DOMAIN 40 303 N-TERMINAL.  
 FT DOMAIN 304 343 EGF-LIKE 1.  
 FT DOMAIN 344 396 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 397 437 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).  
 FT DOMAIN 438 481 EGF-LIKE 4.  
 FT DOMAIN 481 511 TSP TYPE-3 1.  
 FT DOMAIN 511 546 TSP TYPE-3 2.  
 FT DOMAIN 547 569 TSP TYPE-3 3.  
 FT DOMAIN 570 605 TSP TYPE-3 4.  
 FT DOMAIN 606 628 TSP TYPE-3 4.

FT DOMAIN 629 666 TSP TYPE-3 5.  
 FT DOMAIN 667 706 TSP TYPE-3 6.  
 FT DOMAIN 707 742 TSP TYPE-3 7.  
 FT DOMAIN 743 980 C-TERMINAL.  
 FT DISULFID 276 276 INTERCHAIN (PROBABLE).  
 FT DISULFID 279 279 INTERCHAIN (PROBABLE).  
 FT DISULFID 308 319 BY SIMILARITY.  
 FT DISULFID 313 328 BY SIMILARITY.  
 FT DISULFID 331 342 BY SIMILARITY.  
 FT DISULFID 348 359 BY SIMILARITY.  
 FT DISULFID 353 368 BY SIMILARITY.  
 FT DISULFID 371 395 BY SIMILARITY.  
 FT DISULFID 401 412 BY SIMILARITY.  
 FT DISULFID 406 421 BY SIMILARITY.  
 FT DISULFID 424 436 BY SIMILARITY.  
 FT DISULFID 442 456 BY SIMILARITY.  
 FT DISULFID 450 466 BY SIMILARITY.  
 FT DISULFID 468 480 BY SIMILARITY.  
 FT DISULFID 496 501 BY SIMILARITY.  
 FT DISULFID 506 526 BY SIMILARITY.  
 FT DISULFID 542 562 BY SIMILARITY.  
 FT DISULFID 565 585 BY SIMILARITY.  
 FT DISULFID 601 621 BY SIMILARITY.  
 FT DISULFID 624 644 BY SIMILARITY.  
 FT DISULFID 662 682 BY SIMILARITY.  
 FT DISULFID 702 722 BY SIMILARITY.  
 FT DISULFID 738 759 BY SIMILARITY.  
 FT CAROHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CAROHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 980 AA; 108213 MW; 056D41EB6E206FCF CRC64;

Alignment Scores:  
 Pred. No.: 0.000119 Length: 980  
 Score: 144.50 Matches: 29  
 Percent Similarity: 56.52% Conservative: 10  
 Best Local Similarity: 42.03% Mismatches: 27  
 Query Match: 21.50% Indels: 3  
 DB: 1 Gaps: 1

US-09-696-872-24 (1-372) x TSP4\_RAT (1-980)

QY 103 GACCTGGGCGCCGACGATCTTCGGGAACCTGCAGGAACCAACGCGCCTGCAGACGTG 162  
 Db 236 AspPheAenArgGlnPheLeuGlyGlnMetThrGlnLeuAenGlnLeuGlyGlnVal 255  
 QY 163 CGGACCTGCTGCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 222  
 Db 256 LysAspLeuLeuArgGlnGlnValGlnValGlnValGlnValGlnValGlnValGlnVal 275  
 QY 223 TGTGACGGCTGCGGCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 273  
 Db 276 CysGlnAlaCyGlyProLeuLysPheGlnSerProThrProSerThrValAlaPro 295  
 QY 274 CCGCAGCCGAAACCGCAGCCGAAACCG 300  
 Db 296 AlaProProAlaProProThrArgPro 304

RESULT 6  
 CSP\_PLABE STANDARD; PRT; 339 AA.  
 AC P06915;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Circumsporozoite protein precursor (CS).  
 OS Plasmodium berghei.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=5821;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=87089740; PubMed=2432395;  
 RA Eichinger D.J., Arnott D.E., Tam J.P., Nussenzweig V., Enea V.;  
 RT "Circumsporozoite protein of Plasmodium berghei: gene cloning and



RT identification of the immunodominant epitopes.",  
 RL Mol. Cell. Biol. 6:3965-3972(1986).  
 CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
 CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE  
 CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
 CC VERTEBRATE HOST).  
 CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
 CC -1- SIMILARITY: Contains 1 TSP type-1 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M14135; AAA29577.1; -.  
 CC PIR: A44948; OZZOMB.  
 CC InterPro: IPR003067; Crcmpzpzozite.  
 CC InterPro: IPR000884; TSP1.  
 CC Pfam: PF00090; TSP\_1; 1.  
 CC PRINTS: PR01303; CRCMSPRZOITE.  
 CC SMART: SM00209; TSP1; 1.  
 CC PROSITE: PS50092; TSP1; 1.  
 CC Malaria; Sporozoite; Repeat; Signal.  
 CC SIGNAL 1 23 PROBABLE.  
 CC CHAIN 24 339 CIRCUMSPOROZOITE PROTEIN.  
 CC DOMAIN 93 196 13 X 8 AA TANDEN REPEATS.  
 CC DOMAIN 206 238 16 X 2 AA TANDEN REPEATS OF P-Q.  
 CC DOMAIN 266 317 TSP TYPE-1.  
 CC SEQUENCE 339 AA; 37138 MW; E8068A6D11D9551B CRC64;  
 SO  
 Alignment Scores:  
 Pred. No.: 0.00109 Length: 339  
 Score: 133.00 Matches: 20  
 Percent Similarity: 87.50% Conservative: 8  
 Best Local Similarity: 62.50% Mismatches: 4  
 Query Match: 19.79% Indels: 0  
 Gaps: 0  
 DB: 1  
 US-09-696-872-24 (1-372) x CSP\_PLABE (1-339)  
 QY 238 CCGGAGCCGCGAGCCGGAACCGGAGCCGCGAGCCGGAACCGGAGCCGGA 297  
 DB 216 ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 235  
 QY 298 CCGGAGCCGGAAGTACCGGATCATCAGAAAAAGAT 333  
 DB 236 ProGlnProGlnGlyGlyAsnAsnAsnAsnAsnAsn 247  
 RESULT 7  
 ID CSP\_PLABA STANDARD; PRT; 347 AA.  
 AC P23093;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Crcmpzpzozite protein precursor (CS).  
 OS Plasmodium berghei (strain Anka).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5623;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=90221834; PubMed=2183186;  
 RA Lockyer M.J., Davies C.S., Sunbrier A., Sinden R.E.;  
 RT "Nucleotide sequence of the Plasmodium berghei circumsporozone  
 RT protein gene from the ANKA clone 2.34L."  
 RL Nucleic Acids Res. 18:376-376(1990).  
 CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
 CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE

CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
 CC VERTEBRATE HOST).  
 CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
 CC -1- SIMILARITY: Contains 1 TSP type-1 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X17606; CA35608.1; -.  
 CC PIR: S07873; OZZOBK.  
 CC InterPro: IPR003067; Crcmpzpzozite.  
 CC InterPro: IPR000884; TSP1.  
 CC Pfam: PF00090; TSP\_1; 1.  
 CC PRINTS: PR01303; CRCMSPRZOITE.  
 CC SMART: SM00209; TSP1; 1.  
 CC PROSITE: PS50092; TSP1; 1.  
 CC Malaria; Sporozoite; Repeat; Signal.  
 CC SIGNAL 1 23 POTENTIAL.  
 CC CHAIN 24 347 CIRCUMSPOROZOITE PROTEIN.  
 CC DOMAIN 93 204 13 X 8 AA REPEATS.  
 CC DOMAIN 214 247 17 X 2 AA REPEATS OF P-Q.  
 CC DOMAIN 274 325 TSP TYPE-1.  
 CC SEQUENCE 347 AA; 37776 MW; 0EC240EE35681AF8 CRC64;  
 SO  
 Alignment Scores:  
 Pred. No.: 0.00108 Length: 347  
 Score: 133.00 Matches: 20  
 Percent Similarity: 87.50% Conservative: 8  
 Best Local Similarity: 62.50% Mismatches: 4  
 Query Match: 19.79% Indels: 0  
 Gaps: 0  
 DB: 1  
 US-09-696-872-24 (1-372) x CSP\_PLABA (1-347)  
 QY 238 CCGGAGCCGCGAGCCGGAACCGGAGCCGCGAGCCGGAACCGGAGCCGGA 297  
 DB 224 ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 243  
 QY 298 CCGGAGCCGGAAGTACCGGATCATCAGAAAAAGAT 333  
 DB 244 ProGlnProGlnGlyGlyAsnAsnAsnAsnAsnAsn 255  
 RESULT 8  
 ID TEGU\_HSV11 STANDARD; PRT; 3164 AA.  
 AC P10220;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 01-APR-1993 (Rel. 25, Last annotation update)  
 DE Large tegument protein (Vinton protein UL36).  
 OS UL36.  
 OS Herpes simplex virus (type 1 / strain 17).  
 OC Viruses; dsDNA viruses; no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Simplexvirus.  
 OX NCBI\_TaxID=10299;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=88274327; PubMed=2839594;  
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,  
 RA McNab D., Perry L.J., Scott J.E., Taylor P.;  
 RT "The complete DNA sequence of the long unique region in the genome of  
 RT herpes simplex virus type 1."  
 RL J. Gen. Virol. 69:1531-1574(1988).  
 CC -1- FUNCTION: TEGUMENT PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,  
 CC EBV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.



DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Splicing factor, arginine/serine-rich 10 (Putative myelin regulatory  
 factor 1) (MRF-1) (Fragment).  
 DB SFRS10.  
 GN  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=95074156; PubMed=7527040;  
 RA Haque N.S., Buchberg A.M., Khalili K.;  
 RT "Isolation and characterization of MRF-1, a brain-derived DNA-binding  
 protein with a capacity to regulate expression of myelin basic protein  
 gene."  
 RL J. Biol. Chem. 269:31149-31156(1994).  
 CC -1- FUNCTION: CAN BIND TO THE MYELIN BASIC PROTEIN (MBP) GENE MB3  
 REGULATORY REGION AND INCREASE TRANSCRIPTION OF THE MBP PROMOTER  
 IN CELLS DERIVED FROM THE CNS.  
 CC -1- SUBCELLULAR LOCATION: Nucleus (By similarity).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL THE TISSUES EXAMINED (LIVER,  
 KIDNEY, SPLEEN, HEART, LUNG, AND BRAIN).  
 CC -1- PTM: EXTENSIVELY PHOSPHORYLATED ON SERINE RESIDUES IN THE RS  
 DOMAIN (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE SR FAMILY OF SPLICING FACTORS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U14648; AAA64595.1; -.  
 DR MGD; MGI:104846; Sfrs10.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; rrm; 1.  
 KM Nuclear protein; RNA-binding; mRNA splicing; Phosphorylation;  
 KW DNA-binding.  
 FT NON\_TER 1 1  
 FT DOMAIN 167 172 GLY-RICH (HINGE REGION).  
 FT DOMAIN 173 259 ARG/SER-RICH (RS DOMAIN).  
 SQ SEQUENCE 283 AA; 31581 MW; 084EB2C9CC9172AE CRC64;  
 Alignment Scores:  
 Pred. No.: 0.00884 Length: 283  
 Score: 122.00 Matches: 42  
 Percent Similarity: 57.01% Conservative: 19  
 Best Local Similarity: 39.25% Mismatches: 22  
 Query Match: 18.15% Indels: 24  
 DB: 1 Gaps: 6  
 US-09-696-872-24 (1-372) x SFRS\_MOUSE (1-283)  
 Oy 38 TCGCCCTGGCGGCACTGCGAGCGCTGCCAAAAGATCCAGCTGGGTGAGACTGT 97  
 |||||  
 Db 176 SerProAAGAGTGTGCTGCGAGCGCTGCCAAAAGATCCAGCTGGGTGAGACTGT 188  
 |||||  
 Oy 98 GTTCAGACTGCGGCGCGGAGATCTTCGGAACCAACGCGCGCTGACAG 157  
 |||||  
 Db 189 -----TTPSerArg-----ArgLysProSerArgTysSerArg 199  
 |||||  
 Oy 158 ACGTCCGGGAGCTGCTGCGGCGAGGTCAGGAGATCACTTCTGAAAACACGGTGA 217  
 |||||  
 Db 200 Ser-----LysSerArgSer--ArgTlnArgSer--ArgSerArgSer 212  
 |||||  
 Oy 218 TGGAGCTGAGCGGCT--CGCGGCGCGGCGCGAGCCGAAACCGAGCGCGCGAGC 274  
 |||||  
 Db 213 ThrSerLysSerArgSerArgTlnArgTlnArgTlnArgTlnArgTlnArgTln 232  
 |||||  
 Oy 275 CGCAGCCGAACCGGAGCGGAAACCGGAACCGAAGTACCGGATCATCAGAAAAGATG 334

Db 233 ArgSerArgSerArgSerArgSerArgSerArgSerArgSerArgSerArgSerArg 252  
 |||||  
 Oy 335 AGTGTAGCGCGCGCGAGAAAT 355  
 |||||  
 Db 253 SerLysSerArgSerArgSer 259  
 |||||  
 RESULT 11  
 PARR\_TRYB STANDARD; PRT; 129 AA.  
 ID PARR\_TRYB  
 AC P09791;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Procylic form specific polypeptide A-beta precursor (Procylin) (PARR  
 A-beta).  
 GN PARRP-BETA.  
 OS Trypanosoma brucei brucei.  
 OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 OX NCBI\_TaxId=5702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=427;  
 RX MEDLINE=87115776; PubMed=3808022;  
 RA Roditi I., Carrington M., Turner M.;  
 RT "Expression of a polypeptide containing a dipeptide repeat is  
 confined to the insect stage of Trypanosoma brucei."  
 RL Nature 325:272-274(1987).  
 (2)  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=90258895; PubMed=2342468;  
 RX Clayton C.E., Frieri J.P., Itzhaki J.E., Bellofatto V., Sherman D.R.,  
 RA Wisdom G.S., Vijayasarathy S., Mowatt M.R.;  
 RT "Transcription of the procylic acidic repetitive protein genes of  
 Trypanosoma brucei."  
 RL Mol. Cell. Biol. 10:3036-3047(1990).  
 (3)  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=427;  
 RA Vijayasarathy S., Ernest I., Itzhaki J., Sherman D., Mowatt M.R.,  
 RA Michels P.A.M., Clayton C.E.;  
 RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.  
 (4)  
 RP SEQUENCE OF 28-64, AND POST-TRANSLATIONAL MODIFICATIONS.  
 RC STRAIN=427;  
 RX MEDLINE=89359323; PubMed=2475493;  
 RA Clayton C.E., Mowatt M.R.;  
 RT "The procylic acidic repetitive proteins of Trypanosoma brucei.  
 Purification and post-translational modification."  
 RL J. Biol. Chem. 264:15088-15093(1989).  
 CC -1- FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING  
 DIFFERENTIATION IN THE INSECT VECTOR.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X04814; CAA28503.1; -.  
 DR EMBL; M33129; AAA30225.1; -.  
 DR EMBL; X52584; CAA36815.1; -.  
 DR PIR; A26036; A26036.  
 KW Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.  
 FT SIGNAL 1 27  
 FT CHAIN 28 107  
 FT PROPEP 108 129  
 FT DOMAIN 59 106  
 24 X 2 AA TANDEM REPEATS OF [DE]-D.

```

FT CARBOHYD 55 55 O-LINKED.
FT LIPID 107 107 GPI-ANCHOR.
SQ SEQUENCE 129 AA; 13314 MW; 063A1B64D8718B86 CRC64;

Alignment Scores:
Pred. No.: 0.0144 Length: 129
Score: 119.50 Matches: 25
Percent Similarity: 58.82% Conservative: 15
Best Local Similarity: 36.76% Mismatches: 16
Query Match: 17.76% Indels: 12
DB: 1 Gaps: 3

US-09-696-872-24 (1-372) x PARB_TRYBB (1-129)

QY 108 GGGCCCGCAGATGCTTCGGGAACCTGCGAGAACCAACGCGCGCTGACGAGCGCGGA 167
    |||||
DB 30 GYProGIuAaP-----LysGIyLeuThrlYsGIyLysGIyLys----- 43
    |||||

QY 168 CTGGCTGCGCGCAGCGATCAGCGAGATCAGTTCTGAAAAACGCGTGATGAGTGTGA 227
    |||||
DB 44 -----GlyGIyLysGIyThrlYsValSerAspAspAspThrAsnGIyThr-As 59
    |||||

QY 228 CGCGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 287
    |||||
DB 59 PPro---AspProGIuProGIuProGIuProGIuProGIuProGIuProGIuProGIuPr 78
    |||||

QY 288 GCAGCGCGAACCAGCGAACCAGCGA 309
    |||||
DB 78 oGIuProGIuProGIuProGIu 85
    |||||

RESULT 12
TONB_HELPY STANDARD; PRT; 285 AA.
AC 025899;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TONB protein.
GN TONB OR HP1341.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
(1)
RN SEQUENCE FROM N.A.
RP STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Uterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Melman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Frazer C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -!- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
CC MEMBRANE PROTEINS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: ANCHORED TO THE CYTOPLASMIC
CC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; AE000635; AAD08383.1; -.
DR PIR; E64687; E64687.
DR TIGR; HP1341; -.
DR InterPro; IPR003538; TONB.
DR InterPro; IPR006260; TONB_C.
DR Pfam; PF03544; TONB_1.
DR PRINTS; PR01374; TONBPROTEIN.
DR TIGRFAMs; TIGR01352; tonB_Cterm; 1.
KW Transport; Protein transport; Inner membrane; Periplasmic;
KW Transmembrane; Signal-anchor; Repeat; Complete proteome.
FT DOMAIN 1 14 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 15 35 SIGNAL-ANCHOR (POTENTIAL).
FT DOMAIN 36 285 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 285 AA; 31585 MW; 3897A8B7B70BDF8 CRC64;

Alignment Scores:
Pred. No.: 0.0229 Length: 285
Score: 117.00 Matches: 20
Percent Similarity: 79.49% Conservative: 11
Best Local Similarity: 51.28% Mismatches: 7
Query Match: 17.41% Indels: 1
DB: 1 Gaps: 0

US-09-696-872-24 (1-372) x TONB_HELPY (1-285)

QY 238 CCGAGCCGCGAACCAGCGAACCAGCGAACCAGCGAACCAGCGAACCAGCGA 297
    |||||
DB 106 PProGIuProGIuProGIuProGIuProGIuProGIuProGIuProGIuProGIuPro 125
    |||||

QY 298 CCGAACCAGCGAACCAGCGAACCAGCGAACCAGCGAACCAGCGAACCAGCGA 352
    |||||
DB 126 ProGIuPro-LysValGIuGIuValLysGIuGIuGIuProGIuGIuGIuProLys 143
    |||||

RESULT 13
PARB_TRYBB STANDARD; PRT; 115 AA.
AC P14043;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Procyelic form specific polypeptide precursor (Procyelin) (PARP).
GN PROX.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
(1)
RN SEQUENCE FROM N.A.
RP STRAIN=227 (TILPAT);
RX MEDLINE=90067841; PubMed=2573878;
RA Koenig E., Delius H., Carrington M., Williams R.O., Roditi I.;
RT "Duplication and transcription of procyelin genes in Trypanosoma
RT brucei."
RL Nucleic Acids Res. 17:8727-8739(1989).
CC -!- FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING
CC DIFFERENTIATION IN THE INSECT VECTOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; X16015; CAA34147.1; -.

```

DR PIR: S14896; FAUTPC.  
 KW Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.  
 FT SIGNAL 1 27  
 FT CHAIN 28 93 PROCYCLIC FORM SPECIFIC POLYPEPTIDE.  
 FT PROPEP 94 115  
 FT DOMAIN 59 92  
 FT CARBOHD 56 56  
 SQ SEQUENCE 115 AA; 11714 MW; 6D39F0731CF0509E CRC64;  
 Alignment Scores:  
 Pred. No.: 0.0256 Length: 115  
 Score: 116.50 Matches: 22  
 Percent Similarity: 54.41% Conservative: 15  
 Best Local Similarity: 32.35% Mismatches: 17  
 Query Match: 17.34% Indels: 14  
 DB: 1 Gaps: 1  
 US-09-696-872-24 (1-372) x PAR1\_TRYB (1-115)  
 QY 108 GGGCCCGCAGATGCTTCGGGAACCTGCAAGAAACCAACGCGCGCTGACAGACGTGCGGA 167  
 DB 30 GlyProGluuAplysGlyLeuThrlysglyGlylys----- 41  
 QY 168 CTGGCTGGCGACGAGATCAGGAGATCAGTCTCTGAAAAACAGGTGATGAGTGTA 227  
 DB 42 -----GlylysglyGlylysglyThrlyValGlyAlaAsp-AspThrAs 56  
 QY 228 CGCGTGGCGCGCGCGACCGCAACCGCAGCGCGCGCGCGCGCGCAACCC 287  
 DB 56 nGlyThrAspProAspProGluProGluProGluProGluProGluProGluProGluPro 76  
 QY 288 GCAGCCGAAACCGAACCGGAA 309  
 DB 76 OGluProGluProGluProGlu 83  
 RESULT 14  
 PAR1\_TRYB STANDARD; PRT; 143 AA.  
 AC P08469;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Procytic form specific polypeptide B1-alpha precursor (Procylin B1-alpha) (PARP).  
 OS Trypanosoma brucei.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 NC NCB1\_TaxID=5702;  
 RX NCBI\_TaxID=5702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Treu 667;  
 RX MEDLINE=88038823; PubMed=3670296;  
 RA Mowatt M.R., Clayton C.E.;  
 RT "Developmental regulation of a novel repetitive protein of Trypanosoma brucei."  
 RL Mol. Cell. Biol. 7:2838-2844(1987).  
 CC -1- FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING DIFFERENTIATION IN THE INSECT VECTOR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M17027; AAA30222.1; -  
 DR PIR: A26918; A26918.  
 KW Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.  
 FT SIGNAL 1 27  
 FT CHAIN 28 121 PROCYCLIC FORM SPECIFIC POLYPEPTIDE B1-

FT PROPEP 122 143 ALPHA.  
 FT DOMAIN 59 120 TANDEM REPEATS OF E-P.  
 SQ SEQUENCE 143 AA; 14881 MW; E760F52DD46D2E54 CRC64;  
 Alignment Scores:  
 Pred. No.: 0.0255 Length: 143  
 Score: 116.50 Matches: 22  
 Percent Similarity: 54.41% Conservative: 15  
 Best Local Similarity: 32.35% Mismatches: 17  
 Query Match: 17.34% Indels: 14  
 DB: 1 Gaps: 1  
 US-09-696-872-24 (1-372) x PAR1\_TRYB (1-143)  
 QY 108 GGGCCCGCAGATGCTTCGGGAACCTGCAAGAAACCAACGCGCGCTGACAGACGTGCGGA 167  
 DB 30 GlyProGluuAplysGlyLeuThrlysglyGlylys----- 41  
 QY 168 CTGGCTGGCGACGAGATCAGGAGATCAGTCTCTGAAAAACAGGTGATGAGTGTA 227  
 DB 42 -----GlylysglyGlylysglyThrlyValGlyAlaAsp-AspThrAs 56  
 QY 228 CGCGTGGCGCGCGCGACCGCAACCGCAGCGCGCGCGCGCGCGCAACCC 287  
 DB 56 nGlyThrAspProAspProGluProGluProGluProGluProGluProGluProGluPro 76  
 QY 288 GCAGCCGAAACCGAACCGGAA 309  
 DB 76 OGluProGluProGluProGlu 83  
 RESULT 15  
 PARC\_TRYB STANDARD; PRT; 145 AA.  
 AC Q06084;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Procytic form specific polypeptide B-alpha precursor (Procylin) (PARP B-alpha) (PSSA-1).  
 GN PARP.  
 OS Trypanosoma brucei.  
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
 NC NCB1\_TaxID=5702;  
 RX NCBI\_TaxID=5702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90258895; PubMed=2342468;  
 RX Clayton C.E., Fuert J.P., Itzhaki J.E., Bellofatto V., Sherman D.R., Wisdom G.S., Vijayaseethy S., Mowatt M.R.;  
 RT "Transcription of the procytic acidic repetitive protein genes of Trypanosoma brucei."  
 RL Mol. Cell. Biol. 10:3036-3047(1990).  
 CC -1- FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING DIFFERENTIATION IN THE INSECT VECTOR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M3130; AAA30226.1; -  
DR EMBL; X52585; CAA36816.1; -  
DR EMBL; L02933; AAA30228.1; -  
DR PIR; A4418; A4418.  
KW Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.  
FT SIGNAL 1 27 BY SIMILARITY.  
FT CHAIN 28 123 PROCYCLIC FORM SPECIFIC POLYPEPTIDE B-  
FT PROPEP 124 145 ALPHA.  
FT DOMAIN 59 122 BY SIMILARITY.  
FT LIPID 123 123 32 X 2 AA TANDEN REPEATS OF [DE]-P.  
SQ SEQUENCE 145 AA; 15107 MW; B6C3B529BD794CE3 CRC64;

Alignment Scores:

Pred. No.:	0.0255	Length:	145
Score:	116.50	Matches:	22
Percent Similarity:	54.41%	Conservative:	15
Best Local Similarity:	32.35%	Mismatches:	17
Query Match:	17.34%	Indels:	14
DB:	1	Gaps:	1

US-09-696-872-24 (1-372) x PARC\_TRYBB (1-145)

QY	108	GGGCGCGCAGATGCTTCGGGAACTGCGAGAAACAAACGGCGCGCTGCAGAGATGCGGGA	167
DB	30	GIYProGIuAspLysGIYleuThrLysGIYLYS-----	41
QY	168	CTGGCTGCGGAGAGATCAAGTCAAGTCTCTGAAAAACAGGTGATGAGTGA	227
DB	42	-----GIYLeuGIYLYSGLYThrLysValGlyAlaAsp-AspThrAs	56
QY	228	CGCGTGGCGCGCGCAGCGCGCAACCGCGCGCGCGCGCGCGCGCGCGGAAAC	287
DB	56	nGIYThrAspProAspProGIuProGIuProGIuProGIuProGIuProGIuProGIuPr	76
QY	288	GCAGCGGAACCGGAAACCGGAA	309
DB	76	OGIuProGIuProGIuProGIu	83

Search completed: August 26, 2003, 16:22:06  
Job time : 21.5 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: August 26, 2003, 16:09:31 ; Search time 75.5 seconds  
(without alignments)  
2542.929 Million cell updates/sec

Title: US-09-696-872-24

Perfect score: 672  
Sequence: 1 aagcttaccatgggaagta.....aattccatcatgcctcgag 372

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USF70\_spool/US09696872/runat\_26082003.150542.8726/app\_query.fasta.1.519  
-DB=SPTRMBL\_23 -OPMT=faetan -SUFFIX=sept -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-DOCLIGN=200 -THR SCORE=sepc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTMT=plco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09696872 @CGN 1.1.172 @runat\_26082003.150542.8726 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEOQUERY -NEG\_SCORES=0 -WAIT -DISPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTRMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriopl:\*  
17: sp\_archaeopl:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	256.5	38.2	130	Q8N2R4	Q8N2R4 homo sapien

2	247.5	36.8	755	6	Q9BG80	Q9BG80 equus caball
3	245.5	36.5	817	4	O14592	O14592 homo sapien
4	238	35.4	755	11	Q9R0G6	Q9R0G6 mus musculus
5	238	35.4	755	11	O8V154	O8V154 mus musculus
6	195.5	29.1	724	4	Q8N4T2	Q8N4T2 homo sapien
7	153	22.8	1319	4	Q94885	Q94885 homo sapien
8	146.5	21.8	963	11	Q9QYS3	Q9QYS3 mus musculus
9	146.5	21.8	963	11	Q921T2	Q921T2 mus musculus
10	133	19.8	272	5	Q25648	Q25648 plasmodium
11	133	19.8	332	5	Q25649	Q25649 plasmodium
12	132	19.6	3164	12	Q69088	Q69088 human herpe
13	130	19.3	874	3	Q96VJ2	Q96VJ2 pneumocysti
14	129.5	19.3	164	10	Q8LJL1	Q8LJL1 saccharum h
15	128	19.0	277	5	Q27045	Q27045 theileria p
16	127	18.9	2387	16	Q8FPF8	Q8FPF8 escherichia
17	126.5	18.8	230	10	Q8SA70	Q8SA70 saccharum h
18	126.5	18.8	416	5	Q03752	Q03752 plasmodium
19	124	18.5	220	11	Q8R2S3	Q8R2S3 mus musculus
20	124	18.5	220	11	Q8C550	Q8C550 mus musculus
21	124	18.5	910	11	Q9JLE8	Q9JLE8 mus musculus
22	124	18.5	912	11	Q9JLE7	Q9JLE7 mus musculus
23	124	18.5	1494	11	Q88902	Q88902 rattus norv
24	123	18.3	145	5	Q99356	Q99356 trypanosoma
25	123	18.3	675	11	Q9EPT3	Q9EPT3 mus musculus
26	123	18.3	704	11	Q9JLQ4	Q9JLQ4 mus musculus
27	123	18.3	823	11	Q9EOP1	Q9EOP1 rattus norv
28	123	18.3	989	11	Q9JLE9	Q9JLE9 rattus norv
29	122.5	18.2	565	5	O15754	O15754 dictyosteli
30	122	18.2	259	4	Q8WV24	Q8WV24 homo sapien
31	122	18.2	374	2	Q33899	Q33899 streptococc
32	122	18.2	376	2	O54071	O54071 streptococc
33	122	18.2	377	2	Q9S6L9	Q9S6L9 streptococc
34	122	18.2	378	2	Q9L809	Q9L809 vibrio para
35	122	18.2	400	4	O15184	O15184 homo sapien
36	121.5	18.1	170	10	Q9MSX3	Q9MSX3 oryza sativ
37	121	18.0	236	10	Q9SNW2	Q9SNW2 entolophon
38	121	18.0	378	10	Q9ZNV1	Q9ZNV1 zea mays (m
39	121	18.0	125	11	O35745	O35745 mus musculus
40	120.5	17.9	667	5	Q95PJ1	Q95PJ1 trypanosoma
41	120	17.9	273	10	Q9ARY7	Q9ARY7 oryza sativ
42	119.5	17.8	135	5	Q95NM2	Q95NM2 trypanosoma
43	119.5	17.8	139	5	Q95PJ3	Q95PJ3 trypanosoma
44	119	17.7	104	12	Q8V9Y2	Q8V9Y2 choristoneu
45	119	17.7	165	5	Q22168	Q22168 caenorhabdi

# ALIGNMENTS

RESULT 1  
Q8N2R4 PRELIMINARY; PRT; 130 AA.  
AC Q8N2R4  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Hypothetical protein FLJ90027.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryo;  
RA Itoigai T., Oca T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,  
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,  
RA Masuko Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,  
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;  
RT "NEBO human cDNA sequencing project.";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK074508; BAC11031.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 130 AA; 13832 MW; BFE7E5EBB7BACDD CRC64;





Db 10 LeuLeuThrLeuAlaAlaLeuGlyAlaSerGlyInGlyInserProLeuGly----- 27  
QY 94 TGTGTTGAGACCTGGGCGCCGAGATGCTTGGGGAACCTGCAGAAACCAAGCGCGCTG 153  
Db 28 -----SerAspLeuGlyProGlnMetLeuArgGlnLeuGlnThrAsnAlaAlaLeu 45  
QY 154 CAGACGTGCGGCGCTGCGGCGGAGCTGCAGGAGATCACGTTCTGTAATAACAG 213  
Db 46 GlnAspValArgGlnLeuLeuArgGlnInValArgGlnIleThrPheLeuLysAsnThr 65  
QY 214 GTGATGAGTGTACGCGCTGCGGCGCGAG-----CCGACGCG 252  
Db 66 ValMetGlnCysAspAlaCysGlyMetGlnInserValArgThrGlyLeuProSerVal 85  
QY 253 AACCG 258  
Db 86 ArgPro 87  
RESULT 4  
Q9R0G6 PRELIMINARY; PRT; 755 AA.  
ID Q9R0G6  
AC Q9R0G6  
DT 01-MAR-2002 (TrEMBLrel. 13, Created)  
DT 01-MAR-2002 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Cartilage oligomeric matrix protein precursor.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Cartilage.  
RA Fang C., Carlson C.S., Leslie M.P., Tull H., Stolerman E., Perris R.,  
RA Ni L., Di Cesare P.E.;  
RT "Molecular Cloning, Sequencing, Tissue and Developmental Expression of  
RT Mouse Cartilage Oligomeric Matrix Protein (COMP).",  
J. Orthop. Res. 0:0-0(1999).  
DR EMBL; AF033530; AAD01972.1; -.  
DR HSSP; P35444; LVDF.  
DR MGD; MGI:88469; Comp.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR006209; EGF\_Like.  
DR InterPro; IPR00367; tsp\_3.  
DR Pfam; PF02412; tsp\_3; 9.  
DR SMART; SM00179; EGF\_CA; 2.  
DR PROSITE; PS00018; EF\_HAND; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 2.  
KW EGF-like domain; Matrix protein; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 755  
FT SEQUENCE 755 AA; 82352 MW; 7DDFCF443589A0B7 CRC64;  
ALIGNMENT SCORES:  
Pred. No.: 9.23e-14 Length: 755  
Score: 238.00 Matches: 54  
Percent Similarity: 66.32% Conservative: 9  
Best Local Similarity: 56.84% Mismatches: 28  
Query Match: 35.42% Indels: 4  
Gaps: 1  
US-09-696-872-24 (1-372) x Q9R0G6 (1-755)  
QY 10 ATGGAAGTACATGATTAGGCTTGGCGGAGCTGCAGCGCTGCCAA 69  
Db 1 MetGlyProThrAlaCysValLeuValLeuAlaAlaLeuArgAlaThrGlyIn 20  
QY 70 AAGAGATCCAGCTGGGTGAGACTGTTGTCAGACTGGGCGCCAGATGCTTGGGAA 129  
Db 21 GlyInIleProLeuGlyGly-----AspLeuAlaProGlnMetLeuArgGln 36

QY 130 CTGAGAGAAACCAAGCGGCGCTGCAGACGTGCGGAGCTGCAGCGAGCTGCAG 189  
Db 37 LeuGlnGlnThrAsnAlaAlaLeuGlnAspValArgGlnLeuLeuArgHsGlnValLys 56  
QY 190 GAGATCAGTTCCTGAATAACACCGGTATGAGTGTGACGCGGTGCGGCGCGAGCGCGAG 249  
Db 57 GlnIleThrPheLeuLysAsnThrValMetGlnCysAspAlaCysGlyMetGlnProAla 76  
QY 250 CCGAAACCGCAGCGCGAGCGCGAGCGCGAGCGCGAAACCGCAGCGG 294  
Db 77 ArgThrProGlyLeuSerValArgProValProLeuCysAlaPro 91  
RESULT 5  
Q8V154 PRELIMINARY; PRT; 755 AA.  
ID Q8V154  
AC Q8V154  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Cartilage oligomeric matrix protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fang C., Leslie M.P., Nord R., Tian H., Di Cesare P.E.;  
RT "Mouse Cartilage Oligomeric Matrix Protein Genomic Sequence.",  
Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF257516; AAL36518.1; -.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR006209; EGF\_Like.  
DR InterPro; IPR00367; tsp\_3.  
DR Pfam; PF02412; tsp\_3; 9.  
DR SMART; SM00179; EGF\_CA; 2.  
DR PROSITE; PS00018; EF\_HAND; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 2.  
KW EGF-like domain; Matrix protein.  
SQ SEQUENCE 755 AA; 82286 MW; 91758FCF789167EB CRC64;  
ALIGNMENT SCORES:  
Pred. No.: 9.23e-14 Length: 755  
Score: 238.00 Matches: 54  
Percent Similarity: 66.32% Conservative: 9  
Best Local Similarity: 56.84% Mismatches: 28  
Query Match: 35.42% Indels: 4  
Gaps: 1  
US-09-696-872-24 (1-372) x Q8V154 (1-755)  
QY 10 ATGGAAGTACATGATTAGGCTTGGCGGAGCTGCAGCGCTGCCAA 69  
Db 1 MetGlyProThrAlaCysValLeuValLeuAlaAlaLeuArgAlaThrGlyIn 20  
QY 70 AAGAGATCCAGCTGGGTGAGACTGTTGTCAGACTGGGCGCCGAGATGCTTGGGAA 129  
Db 21 GlyInIleProLeuGlyGly-----AspLeuAlaProGlnMetLeuArgGln 36  
QY 130 CTGAGAGAAACCAAGCGGCGCTGCAGACGTGCGGAGCTGCAGCGAGCTGCAG 189  
Db 37 LeuGlnGlnThrAsnAlaAlaLeuGlnAspValArgGlnLeuLeuArgHsGlnValLys 56  
QY 190 GAGATCAGTTCCTGAATAACACCGGTATGAGTGTGACGCGGTGCGGCGCGAGCGCGAG 249  
Db 57 GlnIleThrPheLeuLysAsnThrValMetGlnCysAspAlaCysGlyMetGlnProAla 76  
QY 250 CCGAAACCGCAGCGCGAGCGCGAGCGCGAGCGCGAAACCGCAGCGG 294  
Db 77 ArgThrProGlyLeuSerValArgProValProLeuCysAlaPro 91

```
RESULT 6
Q8N4T2 PRELIMINARY; PRT; 724 AA.
ID Q8N4T2
AC Q8N4T2
DT 01-OCT-2002 (TRMBLrel. 22, Created)
DT 01-OCT-2002 (TRMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TRMBLrel. 23, Last annotation update)
DE Similar to cartilage oligomeric matrix protein (pseudoachondroplasia,
epiphyseal dysplasia 1, multiple).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
SQ
SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC033676; AAH33676.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR SMART; SM00181; EGF; 4.
DR SMART; SM00179; EGF_CA; 3.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 2.
KW Matrix Protein.
SQ SEQUENCE 724 AA; 79696 MW; 9AE2DB3F88815FA1 CRC64;

Alignment Scores:
Pred. No.: 8.12e-10 Length: 724
Score: 195.50 Matches: 41
Percent Similarity: 79.63% Conservatve: 2
Best Local Similarity: 75.93% Mismatches: 4
Query Match: 29.09% Indels: 7
DB: Gaps: 1

US-09-696-872-24 (1-372) x Q8N4T2 (1-724)
QY 118 ATGCTTCGGGAACCTGCGAGAAACCAACGCGCGCTGCGAGAGCTGCGGACTGCTGCGG 177
DB 1 MetleuAAGGluLeuGlnGluThrAsnAlaAlaLeuGlnAspValArgGluLeuLeuArg 20
QY 178 CAGCAGGTGAGGAGATGCTCTGTAAGAAACAGGAGATGAGTGAAGCGCGGCGG 237
DB 21 GlnGlnValArgGluLeuThrPheLeuLysAsnThrValMetGluCysAspAlaCysGly 40
QY 238 CCGCAG-----CCGCAGCCGAAACCG 258
DB 41 MetGlnGlnSerValArgThrGlyLeuProSerValArgPro 54

RESULT 7
Q94885 PRELIMINARY; PRT; 1319 AA.
ID Q94885
AC Q94885;
DT 01-MAY-1999 (TRMBLrel. 10, Created)
DT 01-MAY-1999 (TRMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TRMBLrel. 23, Last annotation update)
DE Hypothetical protein KIAA0790 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
SQ
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
```

```
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
DR EMBL; AB018333; BAA34510.1; -.
DR InterPro; IPR001600; SAM.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00536; SAM; 2.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00454; SAM; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01005; SAM_DOMAIN; 2.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1319 AA; 143613 MW; FB97E509FEDBEP93 CRC64;

Alignment Scores:
Pred. No.: 7.33e-06 Length: 1319
Score: 153.00 Matches: 38
Percent Similarity: 50.47% Conservatve: 16
Best Local Similarity: 35.51% Mismatches: 27
Query Match: 22.77% Indels: 26
DB: Gaps: 5

US-09-696-872-24 (1-372) x Q94885 (1-1319)
QY 31 GCGTTGCTGCGCCCTTGCGGACGCTGCGGCT---GCCAAAAAGATCCAGCTGGGT 87
DB 23 GlyValAlaGlyAlaAlaGlyAlaCysSerAlaGlyAlaArgLeuGlyAlaAlaGly 42
QY 88 GGAAGCTGTTTCTCA-----GACCTGGGCGCCGAGATGCTTGGGAACTG 132
DB 43 GlyAspProAlaSerGlyGlnAlaAlaArgGlyCysGlyAlaAlaArgAlaProArgGlyLeu 62
QY 133 CAGAAACCAACGCGGCGCTGAGAGAGCTGCGGAGCTGCGGCGGACAGTCAAGGAG 192
DB 63 GlyArgThrAlaAlaGlyAla----- 68
QY 193 ATCAGCTTCTGTAAGAAACACGCGTATGAGG--TGTGACGCGTCCGGGCGGACGCGAG 249
DB 69 -----ArgAspThrAlaMetGluAspAlaGlyAlaGlyProGlyProGly 84
QY 250 CCGAAACCGCAGCGCGGACCGGACCGGACCGGACCGGAAACCGGAAACCGGAA 309
DB 85 ProGluProGluProGluProGluProGluProGluProGluProGluProGluProGly 104
QY 310 ---GGTACCGGATCATCAGAA 327
DB 105 ProGlyAlaGlyThrSerGlu 111

RESULT 8
Q9QYS3 PRELIMINARY; PRT; 863 AA.
ID Q9QYS3
AC Q9QYS3;
DT 01-MAY-2000 (TRMBLrel. 13, Created)
DT 01-MAY-2000 (TRMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRMBLrel. 23, Last annotation update)
DE Thrombospondin 4 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
SQ
SEQUENCE FROM N.A.
RX MEDLINE=99431670; PubMed=10501972;
RA Newton G., Wieremowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,
Copeland N.G., Lawler J.;
RT "The thrombospondin-4 gene.";
RL Mamm. Genome 10:1010-1016(1999).
DR EMBL; AF152393; AA032714.1; -.
DR EMBL; AF152392; AAD32714.1; JOINED.
DR HSSP; P35444; IYDF.
MGD; MGI:1101779; Thbs4.
```

DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR006209; EGF\_1like.  
DR InterPro; IPR001129; TSPN.  
DR InterPro; IPR003367; TSPN.  
DR Pfam; PF00008; EGF; 1.  
DR Pfam; PF02210; TSPN; 1.  
DR Pfam; PF02412; TSPN; 9.  
DR SMART; SM00179; EGF\_CA; 2.  
DR SMART; SM00210; TSPN; 1.  
DR PROSITE; PS00018; EF\_HAND; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 2.  
KW EGF-like domain.  
FT NON TER 1  
SQ SEQUENCE 863 AA; 95339 MW; 68E3E8646728E4D CRC64;  
  
Alignment Scores:  
Pred. No.: 2,896-05 Length: 863  
Score: 146.50 Matches: 30  
Percent Similarity: 51.85% Conservative: 12  
Best Local Similarity: 37.04% Mismatches: 36  
Query Match: 21.80% Indels: 3  
DB: 11 Gaps: 1  
  
US-09-696-872-24 (1-372) x Q9Y53 (1-863)  
  
QY 67 AAAAAAGATCCAGCTGGGTGAGACTGTTCTCAGACTGGGCCCGCAGATGCTTCGG 126  
DB : : : : :  
108 GlnGlnSerGluProLeuAlaAlaThrSerThrGlyAspPheAsnArgGlnPheLeuGly 127  
QY 127 GAATGTCAGAAACCAACGCGCGCTGACGACGTCGCGGAGCTGCGGAGCAGGTC 186  
DB : : : : :  
128 GlnMetThrGlnLeuAsnGlnLeuLeuGlyGluValIysAspLeuLeuArgGlnGlnVal 147  
QY 187 AGGAGATCAGCTTCCTGAAACACGCGTGTGATGAGTGTGACGCGTGGCGGCGCAGCG 246  
DB : : : : :  
148 LysGlnThrSerPheLeuArgAsnThrIleAlaGlnCysGlnAlaCysGlyProLeuSer 167  
QY 247 CAGCGGAAACCGCAGCGCAGC-----CCGACGCCGACCGGAAACCGCAGCGGAAA 297  
DB : : : : :  
168 PheGlnSerProThrProAsnThrLeuValProIleAlaProProAlaProProThrArg 187  
QY 298 CCG 300  
DB : : : : :  
188 Pro 188  
  
RESULT 9  
Q9Z1T2 PRELIMINARY; PRT; 963 AA.  
AC Q9Z1T2:  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Thrombospondin-4.  
GN THBS4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99431670; PubMed=10501972;  
RA Newton G., Metemowicz S., Morton C.C., Jenkins N.A., Gilbert D.J.,  
RA Copeland N.G., Lawler J.;  
RT "The thrombospondin-4 gene."  
RL Mamm. Genome 10:1010-1016(1999).  
DR EMBL; AF102887; AAC73003.1; -  
DR HSSP; P35444; 1VDF.  
DR MGD; MGI:1101779; Thbs4.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR006209; EGF\_1like.

DR InterPro; IPR003129; TSPN.  
DR InterPro; IPR003367; TSPN.  
DR Pfam; PF00008; EGF; 1.  
DR Pfam; PF02210; TSPN; 1.  
DR Pfam; PF02412; TSPN; 9.  
DR SMART; SM00179; EGF\_CA; 2.  
DR SMART; SM00210; TSPN; 1.  
DR PROSITE; PS00018; EF\_HAND; 1.  
DR PROSITE; PS01186; EGF\_2; 1.  
DR PROSITE; PS01187; EGF\_CA; 2.  
KW EGF-like domain.  
FT NON TER 1  
SQ SEQUENCE 963 AA; 106366 MW; B8BA83B84F489FB1 CRC64;  
  
Alignment Scores:  
Pred. No.: 2,96-05 Length: 963  
Score: 146.50 Matches: 30  
Percent Similarity: 51.85% Conservative: 12  
Best Local Similarity: 37.04% Mismatches: 36  
Query Match: 21.80% Indels: 3  
DB: 11 Gaps: 1  
  
US-09-696-872-24 (1-372) x Q9Z1T2 (1-963)  
  
QY 67 AAAAAAGATCCAGCTGGGTGAGACTGTTCTCAGACTGGGCCCGCAGATGCTTCGG 126  
DB : : : : :  
208 GlnGlnSerGluProLeuAlaAlaThrSerThrGlyAspPheAsnArgGlnPheLeuGly 227  
QY 127 GAATGTCAGAAACCAACGCGCGCTGACGACGTCGCGGAGCTGCGGAGCAGGTC 186  
DB : : : : :  
228 GlnMetThrGlnLeuAsnGlnLeuLeuGlyGluValIysAspLeuLeuArgGlnGlnVal 247  
QY 187 AGGAGATCAGCTTCCTGAAACACGCGTGTGATGAGTGTGACGCGTGGCGGCGCAGCG 246  
DB : : : : :  
248 LysGlnThrSerPheLeuArgAsnThrIleAlaGlnCysGlnAlaCysGlyProLeuSer 267  
QY 247 CAGCGGAAACCGCAGCGCAGC-----CCGACGCCGACCGGAAACCGCAGCGGAAA 297  
DB : : : : :  
268 PheGlnSerProThrProAsnThrLeuValProIleAlaProProAlaProProThrArg 287  
QY 298 CCG 300  
DB : : : : :  
288 Pro 288  
  
RESULT 10  
Q25648 PRELIMINARY; PRT; 272 AA.  
AC Q25648:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Circumsporozoite (CS) protein (Fragment).  
OS Plasmodium berghei.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxId=5821;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87218962; PubMed=3556207;  
RA Weber J.L., Egan J.B., Lyon J.A., Wirtz R.A., Charoenvit Y.,  
RA Maloy W.L., Hockmeyer W.T.;  
RT "Plasmodium berghei: Cloning of the circumsporozoite protein gene."  
RL Exp. Parasitol. 63:295-300(1987).  
DR EMBL; M25445; AAA29531.1; -  
DR InterPro; IPR003067; Circmepizote.  
DR InterPro; IPR002965; P\_flich\_extensn.  
DR InterPro; IPR000884; TSP1.  
DR Pfam; PF00090; TSP\_1; 1.  
DR PRINTS; PR01303; CRCMSPRZOITE.  
DR PRINTS; PR01217; PRICHEXTENS.  
DR SMART; SM00209; TSP1; 1.  
DR PROSITE; PS30092; TSP1; 1.  
FT NON TER 1  
SQ SEQUENCE 272 AA; 29408 MW; 4FF07FA62B32A051 CRC64;





**This Page Blank (uspio)**